The Bell System Technical Journal

Vol. XIX *October, 1940* *No. 4*

The Carrier Nature of Speech

By HOMER DUDLEY

Speech synthesizing is here discussed in the terminology of carrier circuits. The speaker is pictured as a sort of radio broadcast transmitter with the message to be sent out originating in the studio of the talker's brain and manifesting itself in muscular wave motions in the vocal tract. Although these motions contain the message, they are inaudible because they occur at syllabic rates. An audible sound is needed to pass the message into the listener's ear. This is provided by the carrier in the form of a group of higher frequency waves in the audible range set up by oscillatory action at the vocal cords or elsewhere in the vocal tract. These carrier waves either in their generation or their transmission are modulated by the message waves to form the speech waves. As the speech waves contain the message information on an audible carrier they are adapted to broadcast reception by receiving sets in the form of listeners' ears. The message is then recovered by the listeners' minds.

SPEECH is like a radio wave in that information is transmitted over a suitably chosen carrier. In fact the modern radio broadcast system is but an electrical analogue of man's acoustic broadcast system supplied by nature. Communication by speech consists in a sending by one mind and the receiving by another of a succession of phonetic symbols with some emotional content added. Such material of itself changes gradually at syllabic rates and so is inaudible. Accordingly, an audible sound stream is interposed between the talker's brain and the listener. On this sound stream there is molded an imprint of the message. The listener receives the molded sound stream and unravels the imprinted message.

In the past this carrier nature has been obscured by the complexity of speech.[1] However, in developing electrical speech synthesizers

[1] Speech-making processes are here explained in the terms of the carrier engineer to give a clearer insight into the physical nature of speech. The point of view is essentially that of the philologist who associates a message of tongue and lip positions with each sound he hears. This aspect also underlies the gesture theory of speech by Paget and others and the visible speech ideas of Alexander Melville Bell. The author has been assisted in expressing speech fundamentals in carrier engineering terms by numerous associates in the Bell Telephone Laboratories experienced in carrier circuit theory. Acknowledgment is made in particular of the contributions of Mr. Lloyd Espenschied.

copying the human mechanism in principle, it was soon apparent that carrier circuits were being set up. Tracing the carrier idea back to the voice mechanism there was unfolded, a little at a time, the carrier nature of speech. Ultimately the speech mechanism was revealed in its simplest terms as a mechanical sender of acoustic waves analogous to the electrical sender of electromagnetic waves in the form of the radio transmitter. Each of these senders embodies a modulating device for molding message information on a carrier wave suitable for propagation of energy through a transmission medium between the sending and receiving points.

The Carrier Elements of Speech

This carrier basis of speech will be illustrated by simple speech examples selected to show separately the three carrier elements of speech, namely, the carrier wave, the message wave, and their combining by a modulating mechanism. These illustrations serve the purpose of broad definitions of the carrier elements in speech.

The illustration chosen for the carrier wave of speech is a talker's sustained tone such as the sound "ah." In the idealized case there is no variation of intensity, spectrum or frequency. This carrier then is audible but contains no information, for information is dynamic,[2] ever changing. The carrier provides the connecting link to the listener's ear over which information can be carried. Thus the talker may pass information over this link by starting and stopping in a prearranged code the vocal tone as in imitating a telegraph buzzer. For transmitting information it is necessary to modulate this carrier with the message to be transmitted.

For the second illustration, message waves are produced as muscular motions in the vocal tract of a "silent talker" as he goes through all the vocal effort of talking except that he holds his breath. The message is inaudible because the motions are at slow syllabic rates limited by the relatively sluggish muscular actions in the vocal tract. Nevertheless these motions contain the dynamic speech information as is proved by their interpretation by lip readers to the extent visibility permits. Another method of demonstrating the information content of certain of these motions is the artificial injection of a sound stream into the back of the mouth for a "carrier" whereby intelligible speech

[2] The information referred to is that in the communication of intelligence. There is, however, static information in the carrier itself. This serves for "station identification" in radio and may similarly help in telling whether it was Uncle Bill or Aunt Sue who said "ah."

can be produced from almost any sound stream.[3] The need of an audible "carrier" to transmit this inaudible "message" is obvious.

The final example, to illustrate the modulating mechanism in speech production, is from a person talking in a normal fashion. In this example are present the message and carrier waves of the previous

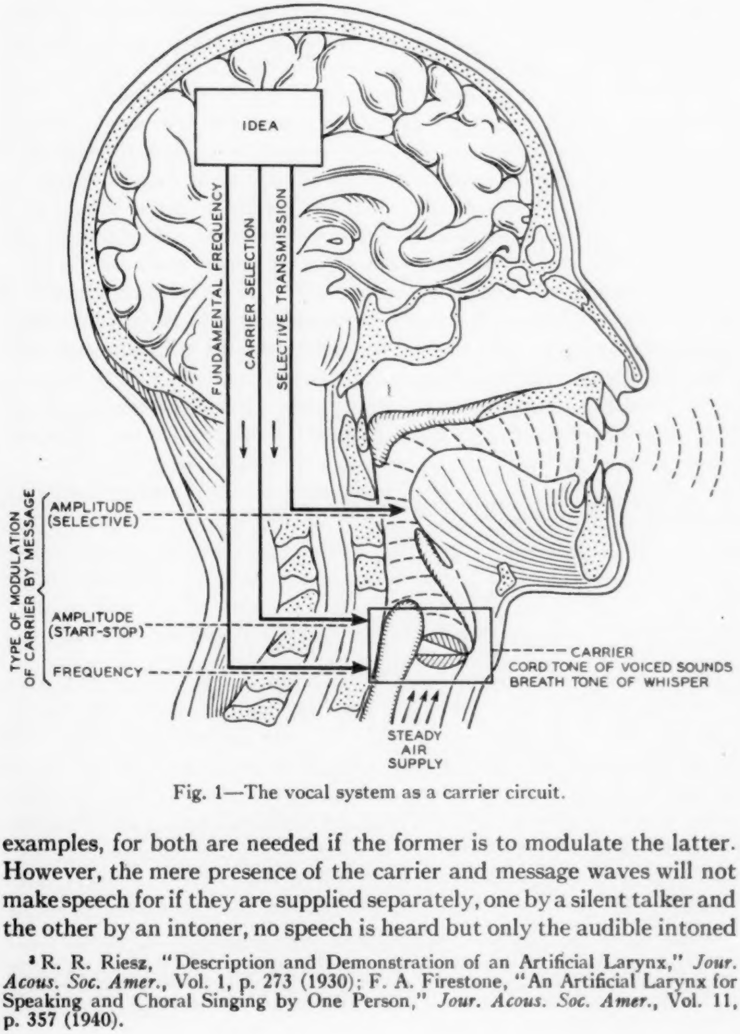


Fig. 1—The vocal system as a carrier circuit.

examples, for both are needed if the former is to modulate the latter. However, the mere presence of the carrier and message waves will not make speech for if they are supplied separately, one by a silent talker and the other by an intoner, no speech is heard but only the audible intoned

[3] R. R. Riesz, "Description and Demonstration of an Artificial Larynx," *Jour. Acous. Soc. Amer.*, Vol. 1, p. 273 (1930); F. A. Firestone, "An Artificial Larynx for Speaking and Choral Singing by One Person," *Jour. Acous. Soc. Amer.*, Vol. 11, p. 357 (1940).

carrier. Ordinary speech results from a single person producing the message waves and the carrier waves simultaneously in his vocal tract, for then the carrier of speech receives an imprint of the message by modulation.

The Speech Mechanism as a Circuit

The foregoing three illustrations by segregating the basic elements in speech production reveal the underlying principles. The present paper treats of these elements as functioning parts of a circuit. In Fig. 1 is shown a cross-section of the vocal system. The idea to be expressed originates in the talker's brain at the left top. Thence, impulses pass through the nerves to the vocal tract with the complete information of the "message," that is to say, what carrier should be used, what fundamental frequency if the carrier is of the voiced type and what transmission through the vocal tract as a function of frequency. The carrier whether voiced or unvoiced is shown for simplicity as arising at the talker's vocal cords. This carrier is modulated to form speech having the complete message imprinted on it preparatory to radiation from the talker's mouth to the ear of the listener, who recognizes the imprinted message.

In discussing the speech mechanism as a circuit, it is clearer to start with a block schematic. Figure 2 has thus been drawn to sketch the

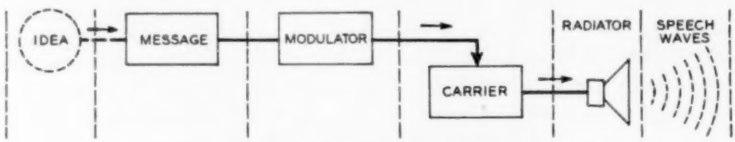


Fig. 2—The basic plan of synthesizing speech.

basic plan of speech synthesizing. As in Fig. 1, the idea gives rise to the message which modulates the voice carrier to produce the speech radiated from the talker's mouth. One can follow the path of the message from its inception in the talker's brain to its radiation from his mouth as an imprint on the issuing sound stream. The progress of the sound stream is also seen from its origin as an oscillatory carrier to its radiation from the talker's mouth carrying the message imprint.[4] The light arrow heads indicate direction of flow while the heavy ones indicate a modulatory control of the carrier by the message. This

[4] Here the carrier path is stressed to show the alteration of the carrier sound stream as it proceeds on its way from the point of origin to the point of radiation. This also accords with the importance of the voice carrier which is received and used by the ear, and thus differs from the treatment of the carrier in simple radio broadcast reception.

modulatory control is exerted on the carrier wave in part as the carrier is generated and in part as it is transmitted after generation.

Relevant Carrier Theory

The heart of the speech-synthesizing circuit of Fig. 2 is the part in which the group of waves making up the message modulate the component waves of the carrier. In any one of these modulations, there is the simple carrier process blocked out in Fig. 3. Here a message [5]

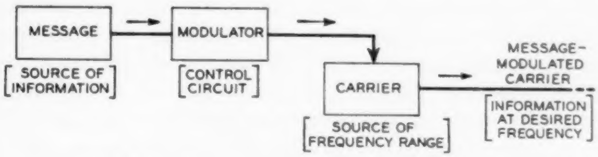


Fig. 3—The elements of a carrier sender.

containing the information modulates a carrier determining the frequency range so that the end product in the form of the message-modulated carrier contains the information of the message translated to frequencies in the neighborhood of the carrier. In this way the carrier sound stream of speech is imprinted with the message.

The prerequisites of the carrier system sender are, as indicated in Fig. 3, first, a carrier wave source; second, a message wave source; and third, a modulating circuit of variable impedance by which the message controls the carrier. The carrier wave is for the simplest case a single sine wave function of time characterized by an amplitude, a frequency and a phase. The message wave as a rule is more complex but may be analyzed as the sum of component sine waves each of which is characterized by its own amplitude, frequency, and phase. In most carrier circuits the frequency range of the message is below that of the carrier. This is true of speech production.

The function of the modulating circuit is supplying a means for the message wave to modify a characteristic of the carrier. If the carrier wave amplitude is modified by the message wave amplitude the process is known as amplitude modulation; if the carrier wave frequency is so modified the process is called frequency modulation while if the carrier wave phase is so modified the process is called phase modulation. No distinction is made as to whether the modification occurs during or

[5] The word "message" has been substituted for the usual carrier term "signal" to avoid confusion since the input signal is commonly speech whereas here the output wave is speech. "Message" seems particularly appropriate with its suggestion of code as in telegraph.

after the generation of the carrier. Modification of the carrier wave characteristics by other than the amplitude of the message need not be considered here. In the voice mechanism significant amplitude and frequency modulations of the carrier occur. Phase modulation takes place also but will not be discussed because the listener's ear is not very sensitive to these phase changes in the carrier.

In attempting to segregate the carrier elements of speech we run into one serious difficulty. In an idealized carrier circuit as shown in Fig. 3 connections can be cut between the two energy sources and the modulator so that each boxed element can be studied independently. With the human flesh of the voice mechanism this is no longer feasible; the use of cadavers would help very little because normal energizing is then impossible. The same difficulty often appears in electrical modulators as, for example, within a modulating vacuum tube where a grid voltage modulates a plate current. In such a case of common parts it is necessary to discuss the action of each of the three elements in the presence of the other two.

With this carrier theory review as a background we are in a position to analyze the three elements making up the carrier transmitting system of the human voice. While the picture presented is over-simplified in details the principles hold and aid in applying carrier methodology to explain the mechanism of speech.

The Voice Carrier

In electrical circuits the carrier is obtained from an oscillatory energy source. The same holds for speech. In the electrical circuit the oscillatory waves (a-c.) are ordinarily generated from a supply of d-c. energy.[6] The same is true in speech with the compressed air in the lungs furnishing the steady supply. Confusion must be avoided, for in speech the conversion of steady to oscillatory energy is often described as *modulation*. Here this conversion of energy form will be considered as an oscillatory action so that the term *modulation* can be reserved for the low-frequency syllabic control of this oscillatory energy to produce the desired speech. *Oscillatory* then will refer to automatic natural responses while *modulatory* will refer to forced responses which are controlled volitionally. This distinction is consistent with carrier terminology.

In the simplest electrical modulating circuits the carrier is a sine

[6] In the usual electrical circuit the carrier is cut off by turning off the output but leaving the carrier oscillator energized as, for example, in voice frequency telegraphy. In the voice mechanism, however, the oscillator is stopped at the source. The difference between the electrical on-off switching and the start-stop switching of speech is not fundamental but results from the use of the most suitable action in each case in view of the conditions prevailing.

wave although this is not true of the damped wave carriers of multi-frequency type once commonly used in spark wave radio telegraphy. The carrier wave in speech is not a simple sine wave. Such a sound would be like a whistle and so too limited for the rich flexibility of speech. Instead the voice carrier is a compound tone having a multiplicity of components of different frequencies which together cover the audible range fairly completely. While these components may be considered as a multiplicity of separate carriers it is simpler to think of the ensemble as a single complex carrier; so this terminology has been used in the earlier carrier illustration and elsewhere in this paper.

Aside from this compound nature of the voice carrier, the voice has two distinct types of carrier, one for voiced and one for unvoiced sounds. Some sounds such as "z" have both types present at the same time but this case may be treated as the superposition of one carrier on the other. For voiced sounds the carrier is the vocal cord tone, an acoustic wave produced by the vibration of the vocal cords consisting of a fundamental frequency component and the upper harmonics thereof. These decrease in amplitude with increasing frequency. For unvoiced sounds the carrier is the breath tone, a complex tone resulting from a constriction formed somewhere in the vocal tract through which the breath is forced turbulently to produce a continuous spectrum of frequency components in the audible range.

These carrier waves must be dissociated from any effects of resonant vocal chambers, for such characterize the speech message rather than the carrier. Furthermore, these carrier waves must be mentally pictured as sustained indefinitely with the starting and stopping of them also characterizing the message wave. Pauses for breath, due to incidental human limitations, do not invalidate the fundamental theory.

The Speech Message

Since a sustained voice carrier has no dynamic flow of information there is need for a source of message waves and a modulating mechanism for imprinting the message on the carrier. Conversely, any variation from the sustained carrier infers the presence of a message wave molding the carrier. The message consists of those articulating, phonating and inflecting motions of the vocal parts which imprint the information on the carrier sound stream. The importance of the message waves cannot be stressed too much. Any impairment of them is an impairment of the message.

The message waves include the motions producing speech changes at infra-syllabic rates, such as the effect of anger when a talker may be high-pitched for many minutes. When the carrier is thus altered over

a long period of time the question arises whether to use a long- or short-term value of the carrier. The answer may well be the same as in the analogous radio problem. If weather causes a carrier frequency to be slightly high all day, this higher value is taken as the normal carrier in studying short-term effects such as the degree of modulation. But in long-term studies of carrier stability the deviations from the mean represent a frequency modulation which is observed as a "message" effect.

Due to the inseparability of the message wave motion and its associated wave of impedance change in the modulating mechanism there may be confusion in distinguishing between the modulating elements and the source of the message waves. The rule followed here is simple. From the standpoint of the human flesh lining the vocal tract, the message source is internal, the modulating elements, external. The message consists of those muscular motions (or pressures or displacements) in the vocal tract which are present in the "silent talker" and are volitional in nature. This definition excludes the oscillatory motions which make up the carrier. The modulating elements are acoustic in nature since the carrier starts as a sound stream and ends as a modulated sound stream.

There are three important variations of the voice carrier and so three types of message and of associated modulation. These variations are: first, selecting the carrier; second, setting the fundamental frequency of the voiced carrier; and third, controlling the selective transmission of the vocal tract.[7] The message waves in the three cases will be discussed with the corresponding modulation reserved for consideration under the next heading.

Selecting the carrier appears as a simple start-stop message, complicated somewhat by the presence of two types of carrier and by locating the constriction for the unvoiced type at several places in the vocal tract. We may think of a start-stop type of message for each point where constrictions are formed, including the vocal cords for the voiced type of carrier. A constriction message may be plotted as the opening between vocal parts at the constriction with critical values for the onset of audible carrier. The constrictions are to a certain extent independent. Thus with the vocal cords vibrating, a constriction from the tongue tip to the upper teeth may also be formed, as in making the "z" sound. Again, in whispering, there may be simul-

[7] A fourth message characteristic prescribes the intensity of the speech. This message may be included in the carrier selection if the carrier is selected for intensity as well as type. The matter of intensity is passed over rather lightly here because a comparison is being developed between the human and electrical speech synthesizers with the final intensity in the latter under control of an amplifier setting.

taneous constrictions, both of the unvoiced type, one at the vocal cords and one in the mouth. As the voice has two distinct types of carrier, the vocal cord tone and the breath tone, the selection sets up one of four carrier conditions at any instant: no carrier, vocal cord tone only, breath tone only, or a combination of vocal cord tone and breath tone. This start-stop message resembles the on-off type of telegraph where switching controlled by other muscular motions sets up speech information in another code, that of telegraph. As mentioned earlier a communication system can be made with the vocal system by starting and stopping a voice carrier in a vocal imitation of a telegraph buzzer. While this would be a clumsy way of communicating information it marks the start-stop control of the voice carrier as a speech message and not part of the voice carrier. Another check is that the "silent talker" does form such constrictions.

The second type of message wave specifies the fundamental frequency with any related voice changes for the voiced type of carrier. This message, in a mechanical form, may be the time variation of the tension of the vocal cords. As the frequency of each upper harmonic is changed in the same ratio as the fundamental frequency, a single parameter suffices for all of the carrier components. The unvoiced carrier has no message of this type impressed since the unvoiced sounds are not characterized by pitch.

The third and final type of message wave controls the selective transmission in the vocal tract. By comparison, the first two types of message are simple, with the selecting of carriers ideally changing all components of the carrier by the same amplitude factor and the fundamental frequency control changing them by a uniform frequency factor. The vocal transmission, however, results from a multi-resonance condition with more than one degree of freedom. There follows a selective amplitude modulation with some carrier components decreasing in amplitude at the same instant that others are increasing. Maximum transmission occurs when a component coincides with an overall resonance, minimum transmission when it coincides with an anti-resonance and intermediate transmission for other cases. The voice message for transmission appears in mechanical form as the displacements of lips, teeth, tongue, etc., with as many such displacements considered as are needed for adequately expressing the speech content. This infers finding the simplest lumped impedance structure equivalent to the distributed impedance structure of the vocal tract to the necessary degree of approximation.

All these mechanical displacements of vocal parts that together constitute the voice message lead to corresponding displacements of

air in the vocal system, resulting in a set of air waves that likewise contain all the information of speech. These airborne message waves, however, are at syllabic rates and so below the frequency range of audibility.

The Voice Modulators

The three voice modulators associated with the three speech messages are the mechanisms of (*a*) selecting the carrier, (*b*) setting the fundamental frequency and (*c*) controlling the selective transmission. The mechanism for starting and stopping a voice carrier is simple. Assume a sustained carrier of either the voiced or unvoiced type. It can be stopped by opening the constriction at which it is formed. This alters the acoustic impedance of the opening which is then the modulating element in this case.

The modulating mechanism for controlling the fundamental frequency appears in the vibrating portions of air at the glottis. The exact mechanism is of no importance here so long as the message wave at the vocal cords finds means for altering the fundamental frequency under the control of the will.[8] This is a case of frequency modulation of multiple carriers harmonically related.

The modulating mechanism for controlling the transmission through the vocal tract as a function of frequency consists of the masses and stiffnesses of air chambers and openings in the vocal tract. These are varied under control of the message in the form of muscular displacements of vocal tract parts. There is a more complicated modulation in the vocal tract than in the usual electrical circuit for amplitude modulation because the varying impedances are reactive in the voice mechanism but resistive in the electrical circuit and also because several independent modulator elements are used in the voice mechanism as against either a single one or a group functioning as a unit in the simple electrical modulator. The reactive nature of the vocal impedances leads to the selective control of the amplitudes of the various harmonics of the voice carrier. The amplitude modulation of each carrier component by the combined message waves produces an output containing the carrier and sideband frequencies.

Comparison of Speech Synthesizing Circuits

The fundamental processes in human speech production are thus analogous to those of electrical carrier circuits. There is a switching of voice carrier energy comparable to that in voice frequency telegraph;

[8] For a simplified theory of the larynx vibration see R. L. Wegel, *Bell Sys. Tech. Jour.*, Vol. 9, p. 207 (1930) and *Jour. Acous. Soc. Amer.*, Vol. 1, Supp. p. 1, April 1930. The analogy of the larynx to a vacuum tube oscillator is described in an abstract, *Jour. Acous. Soc. Amer.*, Vol. 1, p. 33 (1929).

there is an altering of speech frequencies as in frequency modulating circuits; and finally, there is an amplitude modulation to yield a selective transmission of the various carrier components of the voice. However, the voice mechanism differs from the usual carrier circuit markedly as regards complexity. In the voice mechanism there are two types of carrier each with a multiplicity of partial carrier components. The incoming message has a multiple nature. Finally, several modulations take place including both amplitude and frequency types. This multiplicity of carrier relations indicates the wide range of voice phenomena possible.

Any electrical speech synthesizer must be a functional copy of the human speech synthesizer in providing the essential speech characteristics sketched in the preceding paragraph. There have been developed two such electrical synthesizers referred to in the introduction. A brief description of these will be given followed by some circuit comparisons.

These electrical synthesizers are known as the vocoder and the voder. The vocoder was so named because it handles the speech in a coded form; the voder, because it serves as a Voice Operation DEmonstratoR. Considerable interest has been manifested at the public showings of each of these synthesizers, the vocoder in a limited number of lecture demonstrations and the voder at the San Francisco and New York World's Fairs. Circuit details have been published elsewhere.[9]

Of these two speech synthesizers the vocoder was constructed first. It works on the principle of automatically remaking speech under control of spoken speech instantaneously analyzed to derive the code currents for the control. The vocoder as set up for demonstration is shown in Fig. 4.

The voder was derived from the vocoder by substituting manipulative for automatic controls. The resulting voder as displayed at the New York World's Fair is shown in Fig. 5. In the Fair demonstration, repeated continuously at intervals of about five minutes, the male announcer gives a simple running discussion of the circuit with the girl operator replying to his questions by forming sounds on the voder and connecting them into words and sentences. She does this by manipulating fourteen keys with her fingers, a bar with her left wrist and a pedal with her right foot. This requires considerable skill by the operators. The vocoder, automatic in nature, presents no problem of operating technique.

[9] The vocoder in the *Jour. Acous. Soc. Amer.*, Vol. 11, pp. 169–177, October 1939, "Remaking Speech," Dudley; the voder in the *Journal of the Franklin Institute*, Vol. 227, pp. 739–764, June 1939, "A Synthetic Speaker," Dudley, Riesz and Watkins.

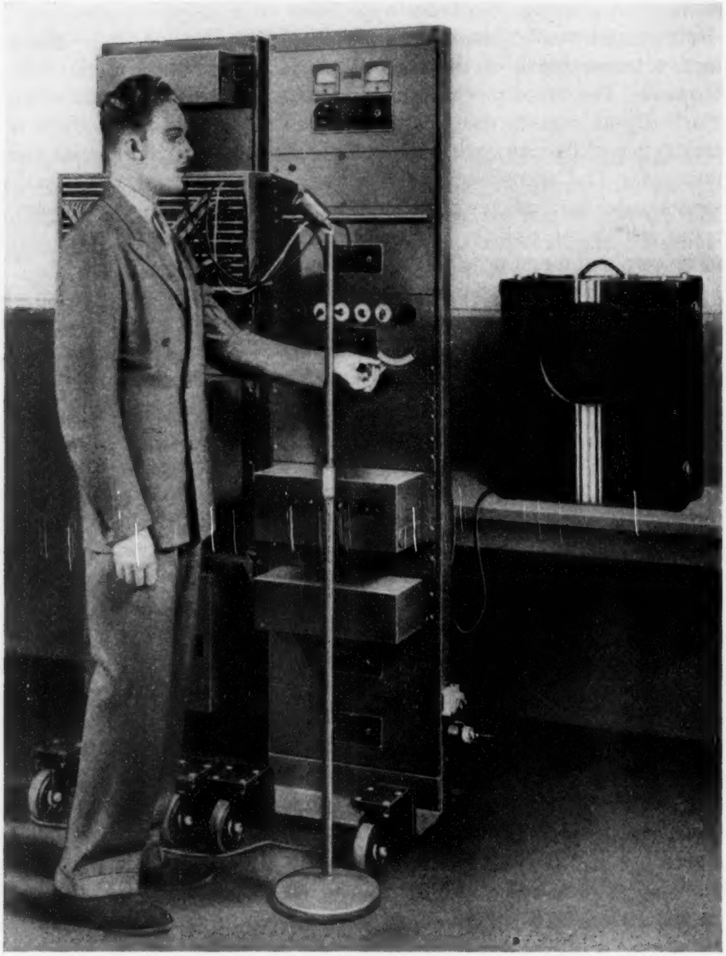


Fig. 4—The vocoder as demonstrated.

Circuit diagrams supply a shorthand for expressing the salient features of electrical circuits. In the next three figures comparative block circuits will be shown for the human and the two electrical speech synthesizers, tracing the communication from the origin of an idea in the communicator's brain to final expression as speech. In each cir-

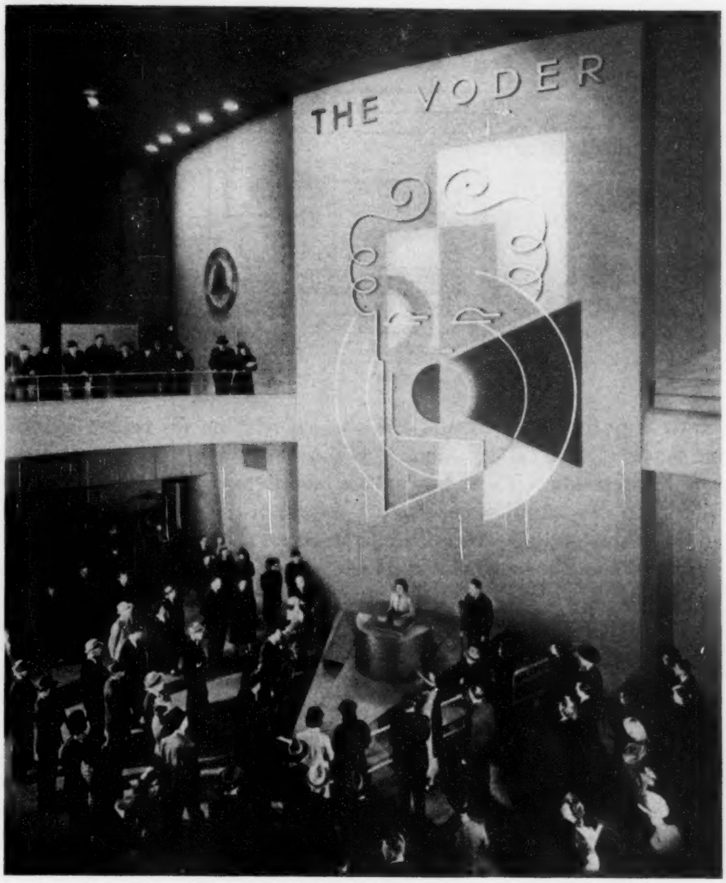


Fig. 5—The voder being demonstrated at the New York World's Fair.

cuit, the arrangement in Fig. 2 will be followed with sufficient detail to show the functional relations of the parts discussed in this paper.

Figure 6 gives a block diagram of the voice mechanism of Fig. 1 with approximating electrical circuit symbols. The same communication paths can be traced. Thus from the talker's brain are sent nerve impulses that set up the message as a set of muscular displacements containing information as to the voice carrier to use, the fundamental frequency for the voiced carrier, and the selective transmission of the vocal tract. The air expelled from the lungs sets up as carriers the

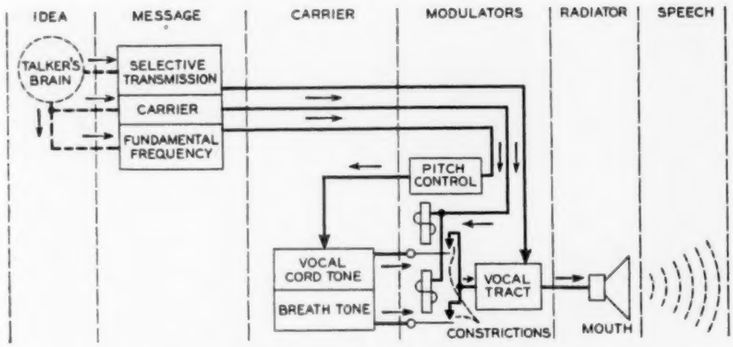


Fig. 6—Block diagram of the voice mechanism.

breath tone for unvoiced and the vocal cord tone for voiced sounds. For simplicity the carrier selection is shown after instead of before the carrier generation. These carriers are modulated by the message wave to produce the output of speech in the form of the message-modulated carrier in the audible range of frequencies.

Figures 7 and 8 show similar block schematics for the vocoder and the voder. The voder circuit has been simplified by the omission of a few controls for easier operation. In these electrical synthesizers, the carrier is provided by a buzzer-like tone from a relaxation oscillator for the voiced sounds and from a hiss-like sound from a gas-filled tube for

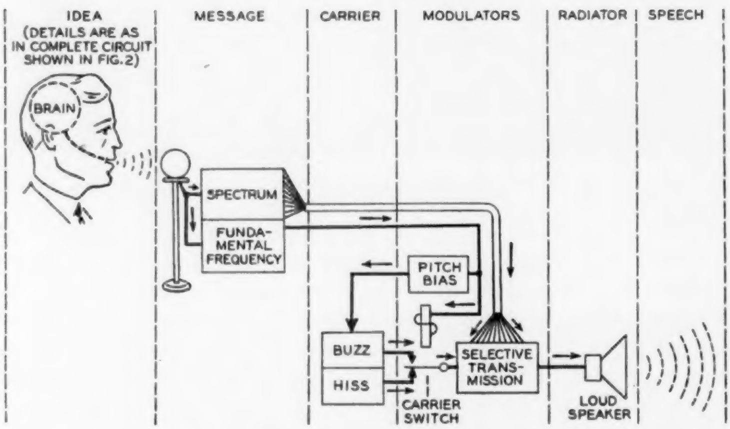


Fig. 7—Schematic circuit of the vocoder.

the unvoiced sounds. In the vocoder, for simplicity's sake, one or the other of these energy sources is used according to whether the sound is voiced or unvoiced, with no provision for the mixed types of sounds found in the human voice. The analyzer of the vocoder derives the

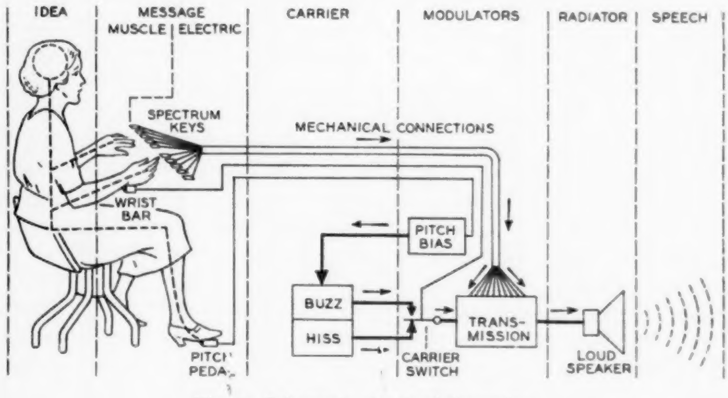


Fig. 8—Schematic circuit of the voder.

original speech message in terms of a modified set of parameters. This analyzer suppresses the original carrier of the talker and so resembles the demodulator in radio reception. The analyzer acts as an electrical ear to tell the artificial vocal system of the vocoder what to say, the whole vocoder acting as a synthetic mimicker.

The basic similarity of the electrical and human speech synthesizers is seen in these figures. In all three cases the message is originated by the brain of the sender of the speech information. There is in each case a transmission of control impulses by the talker's nervous system to the appropriate muscles. The muscles produce displacements of body parts formulating the speech information as a set of mechanical waves. These waves appear in the vocal tract in the case of normal speech; in the fingers, wrist and foot in the case of the voder, but in the case of the vocoder use is made of electrical currents derived from and equivalent to the vocal tract displacements in ordinary speech. In each case the message contains the speech information in syllabic waves. In all cases the message waves control the choice of carrier, the fundamental frequency of the voiced type carrier and the spectrum of power distribution in the speech output. Differences arise in the details rather than in the principles.

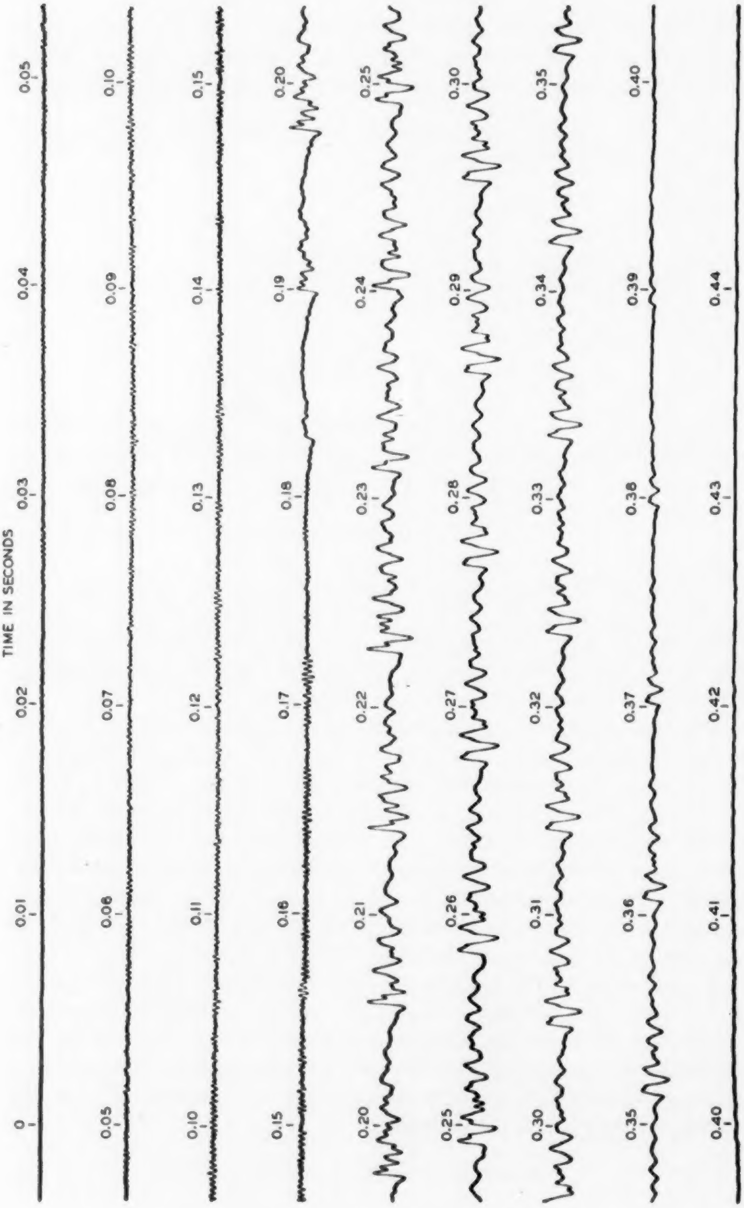


Fig. 9—Oscillogram of the sound "sa."

Speech Characteristics from the Carrier Point of View

Now that the mechanism of speech has been described in carrier terms it is of interest to observe carrier features as they manifest themselves in the characteristics of speech. Some of these can be seen by the eye in speech oscillograms. Some can be demonstrated to the ear with a speech synthesizer such as the vocoder.

For a visual illustration there is shown in Fig. 9 a high quality oscillogram taken from Crandall [10] of the sound "sa" (Plate No. 160. Spoken by M. B.) for a medium-pitched male talker. The carrier shown by the oscillogram is of the unvoiced type for the earlier and of the voiced type for the later part. As one looks at the oscillogram he sees a great mass of the high-frequency components of the carrier. Scrutiny, however, reveals modulated on the carrier the message information in terms of switched energy sources, controlled fundamental frequency and varied transmission characteristic. Shortly after .17 second the switching off of the unvoiced carrier begins. Remnants of the unvoiced carrier can be seen in the voice period just before .19 second and the one starting at about .19 second. The switching on of the voiced carrier appears just after .18 second and seems to be reasonably well completed at the end of the second voice period just before .20 second. This switching was not instantaneous. However, the ear probably does not observe the duration time of the switching. The fundamental frequency falls rapidly at the beginning followed by a leveling out and then a final slight fall in the last few periods. It starts at 140 cycles per second, dropping to around 110 in the level portion, and then to 101 at the end. The resonance conditions cannot be followed too well by eye. However, around .20 second there is a major lower-frequency resonance of about 800 cycles. At .33 second this resonance appears to have increased to 1100 cycles or so. A similar alteration of resonance conditions may be observed if the little shoulder on the rear side of the peak just in front of the .25 second mark is traced in adjacent periods. It can readily be followed back to the third period just before .20 second and can still be seen in the last distinct voicing period starting before .39 second. The dynamic variation of the speech at syllabic rates in accordance with the message content is thus revealed.

For another visual illustration of the speech message Fig. 10 shows a set of oscillograms [11] from the vocoder analyzer for the words "She saw Mary." The oscillogram of the input speech is the trace next to

[10] *Bell Sys. Tech. Jour.*, Vol. 4, p. 586, 1925.

[11] This figure is a copy of Fig. 3 in the paper "The Automatic Synthesis of Speech," Dudley, *Proc. Nat. Acad. Sci.*, Vol. 25, pp. 377–383, July 1939.

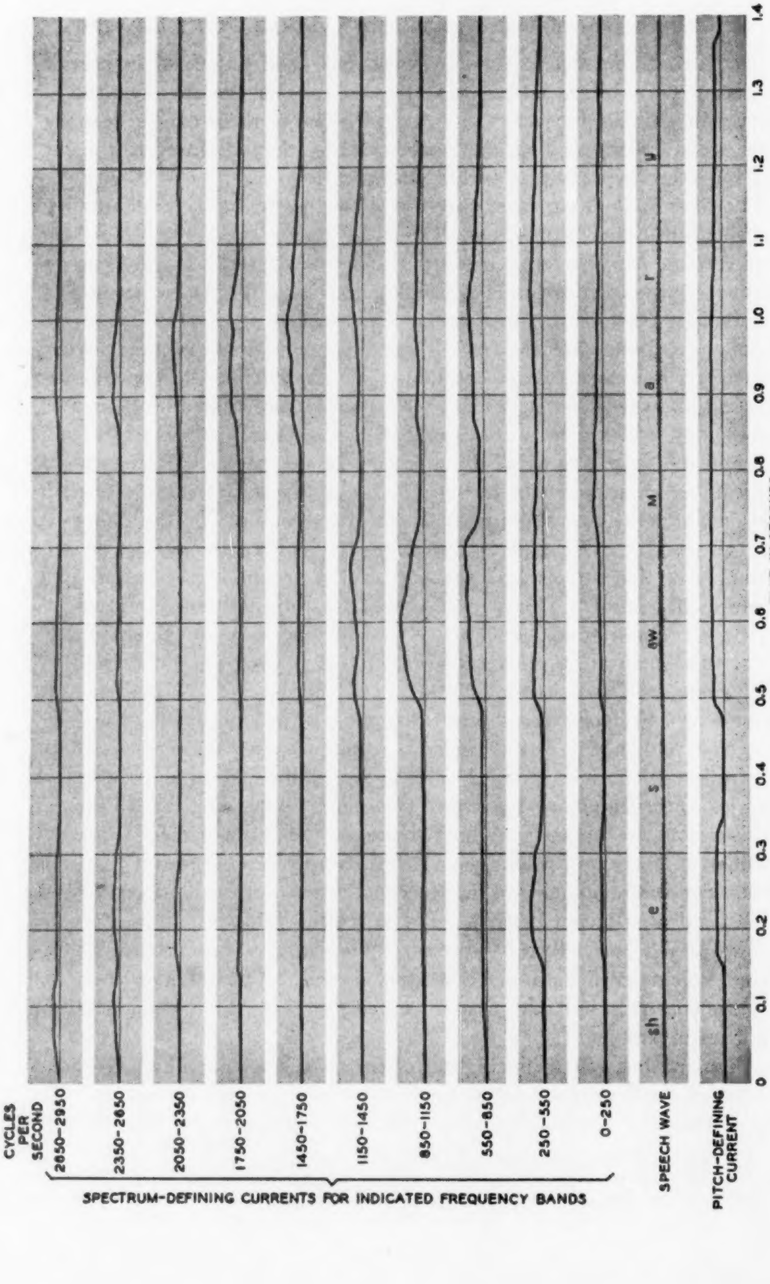


Fig. 10—Coded speech (derived message currents) for the words "She saw Mary."

the bottom. The trace below shows the defining current for the fundamental frequency, while the ten traces above show currents indicating the rectified power in ten frequency bands of 300 cycles width except that the lowest one extends from 0 to 250 cycles. The slow rates of change are noted in the message currents when compared to the original speech wave.

Demonstrations of the vocoder indicate to the ear the carrier nature of speech. Thus the carrier used for remaking speech, whether a monotone or a hiss sound, is observed to have no intelligibility when heard alone. The message currents derived from spoken speech are not audible. However, intelligible "speech" is produced by the modulation of either type of carrier by the message currents of selective transmission. Similarly, there can be used for the carrier a wide variety of sound from the puffs of a locomotive to instrumental music. Upon imprint of the transmission message currents from spoken speech, new forms of odd sounding but nevertheless intelligible "speech" are produced.

The carrier conception of speech reveals what is important and not important in evaluating speech characteristics. An example of interest is the matter of phase. It has long been known that phase was unimportant to the ear at reasonably low listening levels. From the carrier point of view this is natural, for the phase changes referred to are those in the carrier and so, unimportant. When the phases of the message components are altered, there is a very noticeable effect on the ear, for phonetic units are now being shifted.

The great advance in recent years in the application of carrier circuits has been guided by mathematical theory. Since in electrical speech synthesizers the carrier and message currents are separated physically, it is possible to use carrier equations expressing the modulation phenomenon. Similar equations may be written for the voice mechanism as represented by Fig. 6. This has been done in the attached appendix, thus separating speech into syllabic and carrier factors.

APPENDIX

Mathematical Relations

The speech concepts developed in the body of the paper may be expressed in mathematical terms which not only give the fundamental relations in simplest form but also aid in the application of the well-established carrier technique to speech. For voiced sounds, periodic

Fig. 10—Coded speech (derived message currents) for the words "She saw Mary."

by nature, the carrier C_v may be written as a function of the time t thus:

$$C_v = \sum_{k=1}^{n} A_k \cos (kPt + \theta_k). \tag{1}$$

Here C_v is composed of n audible harmonics of relatively high frequencies with the kth of amplitude A_k, frequency kP radians per second, and phase θ_k. The choice of fundamental frequency P is somewhat arbitrary but may well represent the average of the talker over the period of interest.

By modulation processes, there is molded on to this carrier the total message information at the relatively low syllabic frequencies. The message is divided into three parts: (*a*) the starting and stopping of the carrier; (*b*) the instantaneous fundamental frequency; and (*c*) the selective transmission through the resonant vocal tract.[12] These three message functions as they manifest themselves in varying the carrier will be represented by s, p, and r, respectively. Equation (1) will be modified to indicate the effect on the carrier of each of these modulations separately, after which the equation will be rewritten to show the effect of all three acting simultaneously.

The effect of starting and stopping the carrier is described mathematically as a function of time by multiplying C_v by the switching function $s(t)$, giving:

$$\text{Switched } C_v = s(t) \sum_{k=1}^{n} A_k \cos (kPt + \theta_k). \tag{2}$$

For simple on-off switching, $s(t)$ alternately equals zero and unity, although it may in general represent more gradual changes or even any variations of intensity over the frequency range.

The instantaneous fundamental frequency is obtained by multiplying P by the inflecting factor $p(t)$. The effect of the frequency modulation [13] is represented by substituting for Pt the integrated quantity

$$\int_0^t Pp(t)dt = P \int_0^t p(t)dt.$$

Writing this value for Pt in equation (1) gives the inflected carrier wave:

$$\text{Inflected } C_v = \sum_{k=1}^{n} A_k \cos \left[kP \int_0^t p(t)dt + \theta_k \right]. \tag{3}$$

[12] As in the body of the paper, the effect of phase modulation is neglected here.
[13] "Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation," J. R. Carson and T. C. Fry, *Bell Sys. Tech. Jour.*, Vol. 16, p. 513 (1937).

The effect of the selective transmission is allowed for by multiplying C_v by the transmitting factor $r(\omega, t)$, ω indicating that the transmitting factor is a function of frequency at any instant. Applying this factor in equation (1) gives:

$$\text{Transmitted } C_v = \sum_{k=1}^{n} r(\omega, t) A_k \cos (kPt + \theta_k). \tag{4}$$

The r factor is placed inside the summation to indicate that as k changes the different frequencies have different values of the multiplying factor r. If a multiplicity of carrier waves is assumed, the transmitting factor would be $r_k(t)$, individual to the kth component.

In normal voiced speech, S_v, these three modulations are all present simultaneously, so that:

$$S_v = s(t) \sum_{k=1}^{n} r(\omega, t) A_k \cos \left[kP \int_0^t p(t)dt + \theta_k \right]. \tag{5}$$

Equation (5) shows how the message in the form of the s, r, and p functions has imprinted its characteristics on the original carrier C_v of equation (1).

The derivation of (5) was for voiced speech. Unvoiced speech, however, is also covered by (5) as a degenerate case. Nevertheless, further information is presented by writing out the unvoiced carrier separately. For unvoiced speech, the frequency P approaches zero and the number of terms, n, approaches infinity, giving an integral instead of a finite sum of components in equations (1) and (5). The unvoiced carrier C_u is then:

$$C_u = \int_{\omega_1}^{\omega_2} A(\omega) \cos [\omega t + \theta(\omega)] d\omega \tag{1'}$$

and the unvoiced speech:

$$S_u = s(t) \int_{\omega_1}^{\omega} r(\omega, t) A(\omega) \cos [\omega t + \theta(\omega)] d\omega \tag{5'}$$

with the continuously variable frequency ω (radians per second) varying over the audible range of energy contribution from ω_1 to ω_2 and the unvoiced carrier spectrum defined by amplitude $A(\omega)$ and phase $\theta(\omega)$. The unvoiced speech has no inflecting factor but does have switching and transmitting factors to make up the message impressed on the carrier.

Manufacture of Quartz Crystal Filters

By G. K. BURNS

Quartz crystal filters used in modern carrier systems present new problems in manufacturing technique. In the assembly and testing of the filters and in the production of component crystals, coils and condensers, special factory facilities are required for accurate measurement of frequency and control of atmospheric conditions. The manufacture of quartz crystal plates in particular combines several fields of applied science, including crystallography, precision grinding, vacuum technique and high frequency electrical measurement. Inductance coils and fixed and variable condensers for use in crystal filters must consistently meet advanced requirements, especially in regard to stability. The assembly of these components into filters resembles the manufacture of radio receivers, differing mainly because of smaller quantity requirements. Testing equipment must permit rapid shop adjustment and test of the completed filters with laboratory precision.

Introduction

ELECTRICAL wave filters employing quartz crystals [1] are used extensively in broad band carrier systems [2, 3] recently introduced into commercial service. Such crystals exhibit the property of piezoelectricity; that is, an electrical voltage applied to the terminals of a crystal causes a mechanical distortion of the quartz, and vice versa. Because of this interrelation a plate of quartz, at frequencies near its mechanical resonance, behaves electrically like the coil and condenser combination shown in Fig. 1. The series inductance and capacitance

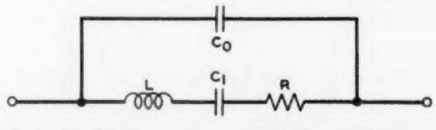


Fig. 1—Equivalent circuit of a quartz crystal plate. Elements L, C_1 and R are associated with the piezo-electric property and mechanical resonance of the crystal, while C_0 represents capacitance between the electrodes.

represent the mass and elasticity of the plate, respectively, while the shunt condenser represents the capacitance between faces of the crystal. The damping of such a plate may be made very low, giving a ratio of reactance to resistance (commonly termed Q) of 15,000 or

[1] Numbered references are listed at end of paper.

higher, as compared with a practical limit of 300 for coils. Stability of resonance frequency and compactness of dimensions are two further respects in which quartz crystals surpass the best coils and condensers available.

Filters designed to utilize these properties generally consist of one or more crystal plates, plus such condensers, inductance coils and resistances as may be required to give the desired overall performance. The principal types used in the Bell System operate at frequencies ranging from 40 to 600 kilocycles and transmit bands varying from 5 cycles to 6 kilocycles in width. Physical dimensions range up to $3 \times 5 \times 16$ inches.

Unusual manufacturing requirements are imposed by the nature of these filters and of the systems in which they are used. Adjusting tolerances and stability requirements, for example, range from $\pm$ 20 to $\pm$ 200 parts per million on crystals and on coil-and-condenser circuits used in crystal filters. Transmission losses must be measured to accuracies of the order of $\pm$ .03 db at 100 KC. To insure stability of adjustment during service life, component apparatus must be protected against dust and excessive humidity. Methods of assembly and testing must be adaptable to a variety of types of filters, one of which, the channel filter,[4] is manufactured by the Western Electric Company in quantities of 1500 to 5000 per year, while the others range from 10 to 1000 per year. Long service life must be assured by proper choice of materials and technique.

In order to satisfy such requirements special manufacturing procedures are necessary. In reviewing these features it will be convenient to consider first those methods or facilities which are used in several or all stages of the manufacture of crystal filters, second the methods employed in producing component apparatus for such filters—particularly crystals, coils and condensers—and finally the technique of assembling and testing the complete filters.

General Facilities

A primary requisite in the adjusting and testing of both crystals and crystal filters is the precise measurement of frequency. The equipment used for this purpose includes a standard frequency generator containing a 100 KC crystal oscillator. This generator normally maintains a frequency accuracy of about 1 part in 2,500,000 operating under the control of the Bell System master frequency standard in New York, but will remain accurate within 1 part in 1,000,000 even though the master signal is interrupted for as much as 24 hours. Three sub-harmonics of 100 KC, namely, 100 c.p.s., 1000 c.p.s. and

10,000 c.p.s., are distributed to all test positions. Oscillators supplying the individual test sets are provided with cathode ray oscilloscopes, by means of which they can be synchronized with any multiple of the three standard frequencies. To set up an odd frequency not coinciding with any multiple it is necessary to interpolate dial readings between two synchronized points.

Control of atmospheric conditions also plays an important part in the manufacture of crystal filters. The temperature coefficient of frequency of the crystals most commonly used is about 15 parts per million per degree Fahrenheit. For some filters, in order to secure uniform performance throughout the temperature and frequency ranges encountered in service, these crystals must be adjusted within tolerances as small as 40 parts per million. Fluctuations of as little as 2° F., in such cases, must be taken into account during the adjustment of the crystals. In addition, crystals, coils and condensers are all sensitive to the effects of excessive humidity. To minimize such difficulties, the assembly and testing of these components and of the filters in which they are used are carried out in air conditioned rooms controlled at 75° ± 2° F. and approximately 40 per cent relative humidity.

Crystals

Of the several component parts used in crystal filters, the first to be considered in detail are logically the quartz crystals themselves. Their properties of low loss and high stability are primarily responsible for the unusual performance of filters in which they are employed.

Natural deposits in the earth constitute the sole source of supply of quartz crystals, since no practical method of producing them synthetically has been developed. "Raw" crystals suitable for use in filter manufacture must be unusually large and free from flaws. The principal source is Brazil, the bulk of the quartz being brought in by native prospectors to trading posts and shipped to this country via Rio de Janerio and other coastal cities. The crystals usually range between 3 and 10 pounds in weight, with occasional pieces reaching 100 pounds.

The raw quartz passes through successive stages of inspection and selection, commencing at the trading post and culminating in careful examinations before and during the cutting operations. A concentrated beam of light from an arc lamp (see Fig. 2) is used in locating internal flaws, which generally appear as small bubbles and inclusions of foreign matter. Quartz takes two distinct forms, left-hand and right-hand, having opposite piezo-electric polarities. Portions of raw crystals containing both forms are not usable. This condition, called

Fig. 2—Inspection of quartz crystals. An arc light beam aids in the detection of internal flaws.

"twinning," appears as shown in Fig. 3 when observed with polarized light.

For use in filters, quartz must be cut into rectangular plates properly oriented with respect to the electrical, mechanical and optical axes of the crystal, as shown in Fig. 4. A polariscope and an X-ray spectroscope are used in locating these axes to an accuracy of $\pm$ 0.25 degree. For the majority of applications the plate is cut in the plane of the mechanical and optical axes, with the long dimension set at an angle of 18.5° from the mechanical axis. This orientation eliminates secondary resonances in the completed crystal and makes the primary resonance frequency relatively independent of slight errors in orientation. For applications requiring a low coefficient of resonance frequency versus

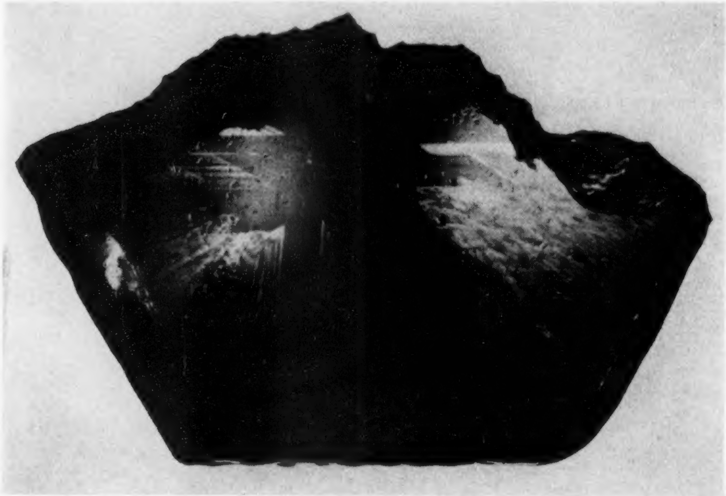


Fig. 3—Right and left-hand twinning in quartz as seen by polarized light.

temperature, plates are cut with their long dimension 5° from the mechanical axis. Tolerances in cutting and grinding to thickness, length and width prior to calibrating are of the order of .001 mm., requiring the use of technique similar to that employed in the manufacture of gage blocks. A few standard thicknesses, ranging from .020 to .060 inch, are used for most crystal plates. Lengths vary from 0.5 to 2.0 inches while widths range from 0.15 to 1.5 inches. Because of unavoidable waste in the cutting and grinding operations and the rejection of quartz containing flaws, only a small portion of the material entering the cutting room finds its way into finished plates.

Up to this point the cutting and grinding are purely mechanical operations, directed toward securing prescribed physical dimensions. During final adjustment and in service, however, the crystal plate must be connected as an electrical element. Electrodes are provided by coating the major surfaces of the plate with aluminum, using a process of evaporation and condensation in a vacuum, similar to that employed in the silvering of telescope mirrors. If the plate is to be used in a balanced filter section which requires a pair of crystal elements of the same frequency, as is frequently the case, the plating on each face is then divided in half along the longitudinal axis. This division, one-hundredth of an inch wide, must have a d.c. insulation resistance of at least 100 megohms to insure proper operation in some types of crystal filters.

Preliminary tuning is accomplished with a fixture, simulating the final holder, which grips the plate at the center by four contact points, one on either side of the division in the plating of each face. These

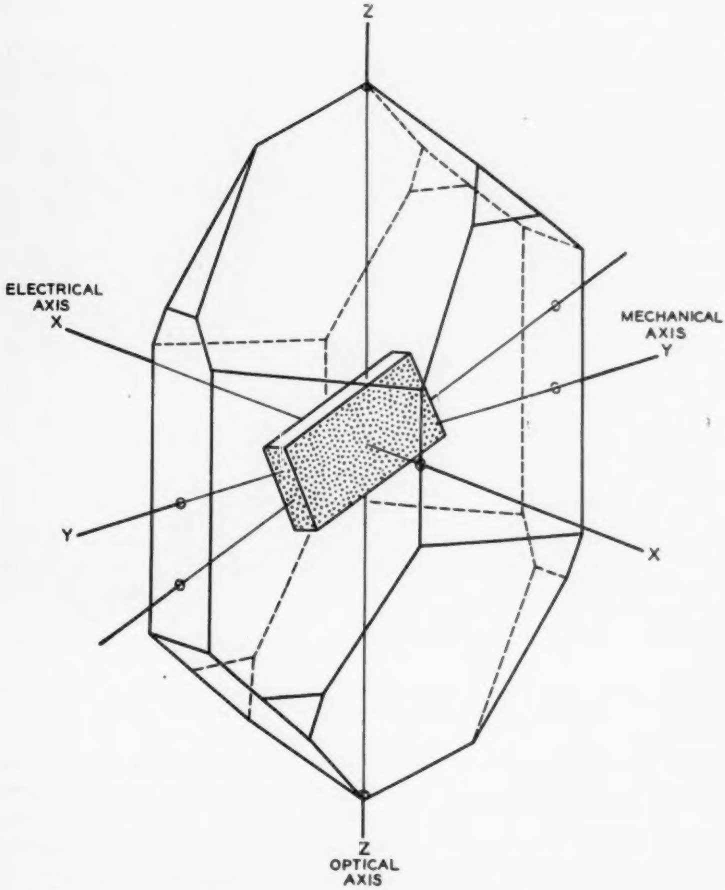


Fig. 4—Orientation of a typical quartz plate with respect to its electrical, mechanical and optical axes.

contacts introduce very little damping, since the mode of vibration normally employed is longitudinal, with maximum amplitude at the ends of the plate and a node at the center. The test set-up normally used consists of two oscillators with a meter arranged to read the

difference between their frequencies. One oscillator is controlled by the crystal plate being tuned and operates at its resonance frequency; the other is controlled by a standard crystal of the desired frequency. Starting 100 to 200 cycles low, the plate is ground on the ends until its frequency approaches that of the standard.

The plate is then transferred from the fixture to its final holder, shown in Fig. 5. This mounting normally accommodates two plates

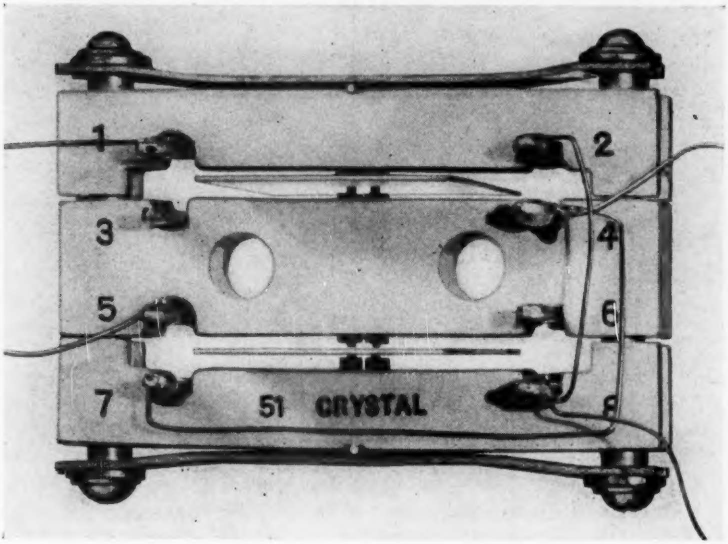


Fig. 5—Crystal plates mounted in holder. The four points at the center of each plate provide electrical contact and mechanical support.

of different frequencies, each supported at its nodal point by contacts projecting from ceramic blocks. The entire assembly is held together by a spring suspension in order to apply uniform pressure at all contacts. To minimize damping, the contacts must be accurately aligned and the quartz plates must be carefully centered upon them.

A final adjustment of frequency is now performed, as shown in Fig. 6. Permissible tolerances vary from $\pm$ 20 to $\pm$ 150 parts per million for different types of crystals. Crystals having the broader tolerances and substantial quantity requirements are adjusted by comparison with a standard crystal, as in the case of preliminary tuning. The test set shown at the left in Fig. 6 is being used for this purpose. The upper and lower panels are the oscillators controlled by

the standard and the test crystals, respectively, while the center panel indicates the frequency difference between them. For very accurate work and for periodic checks of the standard crystals it is necessary to use a precision oscillator, shown at the right.

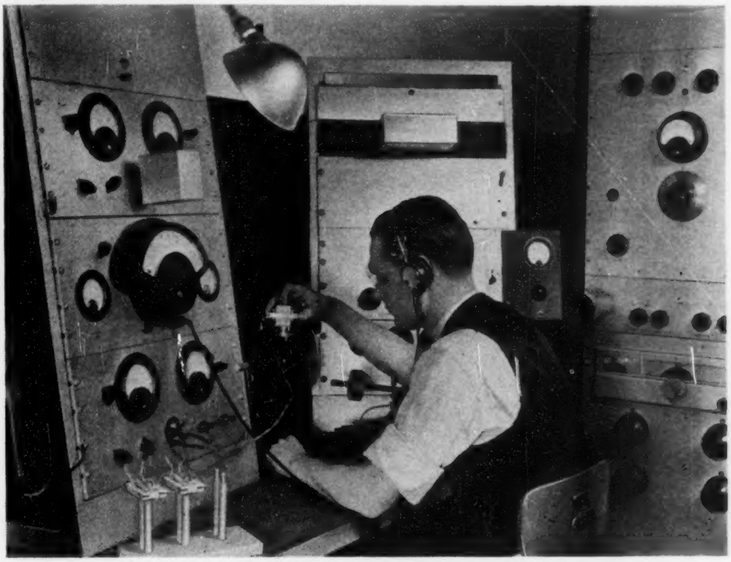


Fig. 6—Final tuning of a crystal plate using a standard crystal (in small box at upper left) for comparison.

Occasionally, in the course of adjustment, plates are carried too high in frequency. In such instances, as a result of the standardization of thicknesses mentioned previously, the plate normally can be salvaged by grinding it to the dimensions of the next higher frequency plate of the same thickness.

Aging occurs in both resonance frequency and effective resistance, as slight strains created in the quartz and in the contacts during adjustment relieve themselves. The greater part of the aging takes place during the first few hours after calibration and nearly all of it during the first week. In general, the frequency rises a few cycles and the resistance drops slightly. Crystals on which the frequency tolerance is approximately equal to the shift due to aging are stabilized by one or more temperature cycles, prior to final measurement of frequency and resistance.

Coils

Inductance coils are used in some crystal filters, particularly the channel filter for the newer types of carrier telephone systems. Since it may be necessary to connect as many as ten such filters in tandem in a long-distance circuit without appreciable impairment of the quality of transmission, the filters must meet exacting requirements not only

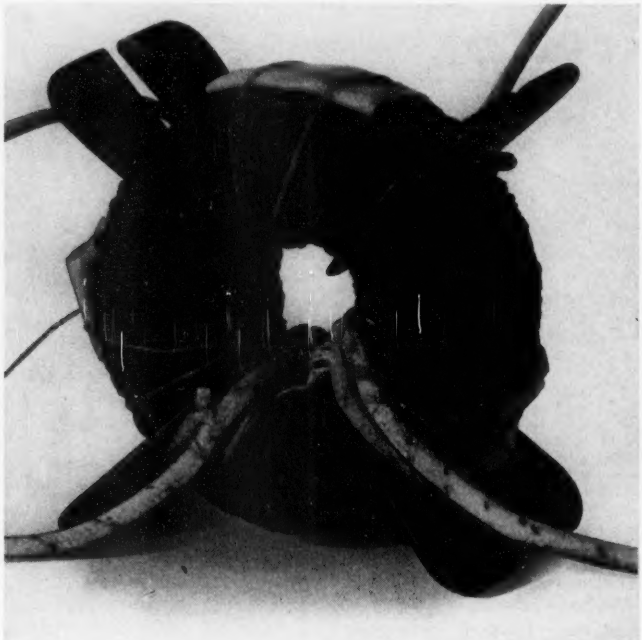


Fig. 7—Toroidal inductance coil used in crystal channel filter, shown before potting.

at the time of their manufacture but throughout service life. Consequently inductance coils used in the filters must exhibit little aging or shift with temperature, either in inductance or in effective resistance. Losses must be kept low in order to meet a Q requirement of approximately 200. The types employed in channel filters range from 25 to 50 millihenries in inductance and from 60 to 120 kilocycles in operating frequency.

Unusual features of design and manufacture are employed to meet these requirements. The coil is essentially a toroidal winding with low distributed capacitance, applied to a permalloy dust core, im-

pregnated and potted in wax. A molded jacket with protruding fins, placed around the core, reduces the capacitance from windings to core and improves the uniformity of the windings. The coils are adjusted to within $\pm$ 1 per cent for inductance and 2 per cent for inductance unbalance by removal of excess turns, all adjustments being made at low frequency. Figure 7 shows a coil at this stage of manufacture.

The coil is then potted in a copper can and a cover soldered in place. Final test simulates actual service conditions. The coil is resonated with an external variable condenser at the operating frequency for which it is designed.

Condensers

Nearly all crystal filters contain condensers shunted across the crystal elements. These condensers must meet stability requirements similar to those already mentioned in connection with coils.

One form of fixed condenser, used where small values of capacitance and high stability are required, is illustrated in Fig. 8. Silver is fused

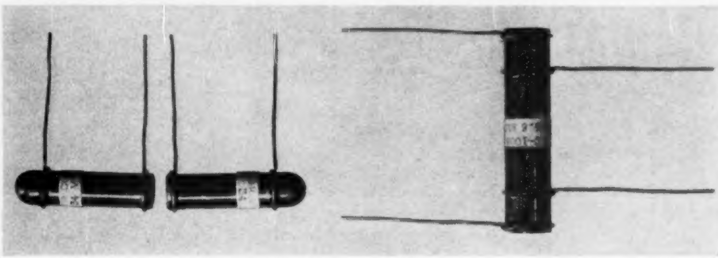


Fig. 8—Silvered glass condensers used in crystal filter applications where high stability is required.

to the inside and outside of a glass tube by applying a coating of silver paste and firing the tube in an oven. A gap is left uncoated on the outer surface near the open end and leads are soldered to the silver on both sides of the gap. The capacitance is then adjusted to the required value, within approximately $\pm$ 1.5 mmf., by scraping off a portion of the silver coating. Capacitances up to 80 mmf. are realized by this means. Two condensers may be combined in a single unit, as shown at the right in Fig. 7. The completed condenser is dipped in varnish to protect the silver from corrosion.

Pairs of such condensers, matched to each other within 0.4 mmf., are required in some types of crystal filters. This precision is achieved by manufacturing a quantity of condensers of the correct nominal capacitance and sorting them into close-limit groups after final measurement.

Variable air condensers are used in adjusting the assembled filter. For the channel filter four such condensers are manufactured on a single ceramic base, as shown in Fig. 9, to eliminate unnecessary parts

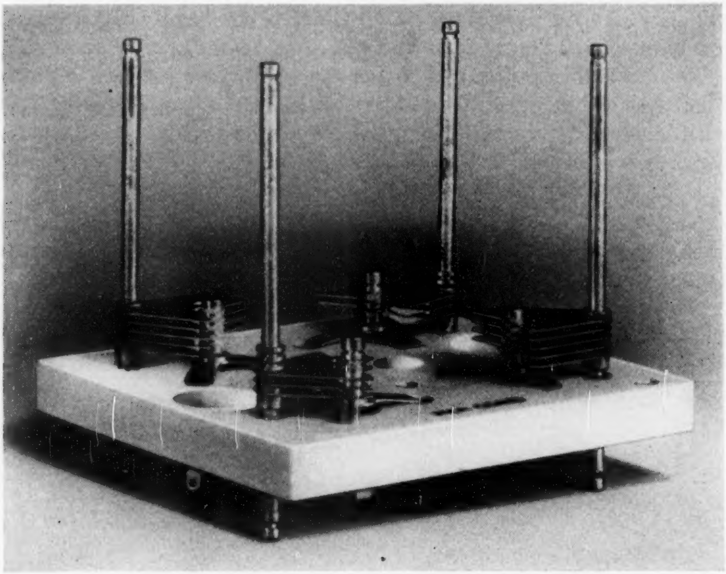


Fig. 9—Four-section variable air condenser.

and reduce assembly cost. The posts supporting the stator plates are extended both upward and downward to serve as convenient terminals for leads from adjacent pieces of apparatus. Freedom from binding is important, since condensers in crystal filters must be adjusted through angles as small as 2 minutes. To insure smooth adjustment the rotor shafts and their bearings are held to close dimensional tolerances and lubricated with petrolatum. Stability is secured by the use of thrust springs providing a substantial holding torque.

Assembly

The foregoing components—crystals, coils and condensers—are assembled into complete filters by methods somewhat similar to those employed in the manufacture of radio receivers, the principal differences arising from smaller volume requirements. The channel filter alone is produced in sufficient quantities (1500 to 5000 per year) to warrant a substantial degree of tooling. Figure 10 illustrates the

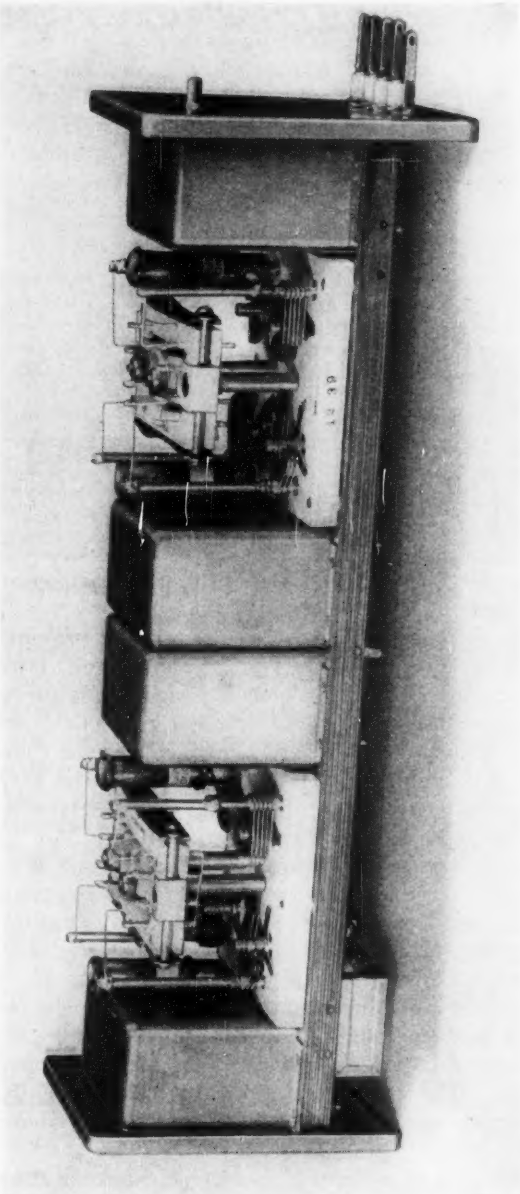


Fig. 10—Internal assembly of crystal channel filter, used in Type *J*, *K* and *L* carrier systems.

internal assembly of this type of filter. The chassis consists of a pair of perforated brass angles running the length of the assembly and spot-welded to a cover at each end. Coils and condensers are riveted to the angles, while the crystal holders are mounted with rubber shock absorbers on studs extending upward from the ceramic bases of the variable condensers. External leads are brought out through copper-to-glass seals to terminals, as shown in Fig. 11, since in final assembly

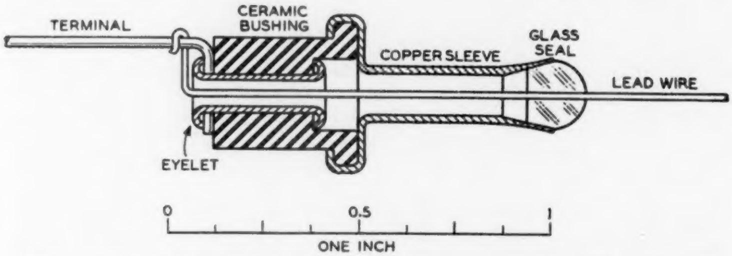


Fig. 11—Terminal used in hermetically sealed filters. The copper sleeve, bonded to the lead wire by means of an insulating glass bead, is soldered into the container of the filter in final assembly.

the filter must be hermetically sealed to protect components from moisture and dust.

In wiring the filter special precautions are taken to prevent foreign materials from being deposited on crystal plates, thereby introducing mechanical damping, and from lodging in variable condensers, where electrical leakage must be avoided. Internal connections are made with bare tinned wire. Rosin flux remaining on soldered connections is washed off with a solvent. Dust and other particles in variable condensers are blown out with air. The filter then undergoes a careful visual inspection and a 500 volt d.c. insulation test.

At this stage it is generally necessary to adjust certain of the filter elements, usually variable condensers, in order to compensate for manufacturing variations in other elements and for parasitic effects such as capacitance of the wiring to ground. A general view of the testing equipment used for this purpose is shown in Fig. 12. One or more of three methods of adjustment are employed, namely, (*a*) transmission loss, (*b*) resonance and (*c*) capacitance. The first and second of these are utilized on the channel filter, the schematic of which is shown in Fig. 13. The two sections are adjusted independently before the resistance pad, seen at the center of the figure, is inserted to connect them.

In transmission loss adjustment of the channel filter, the attenuation

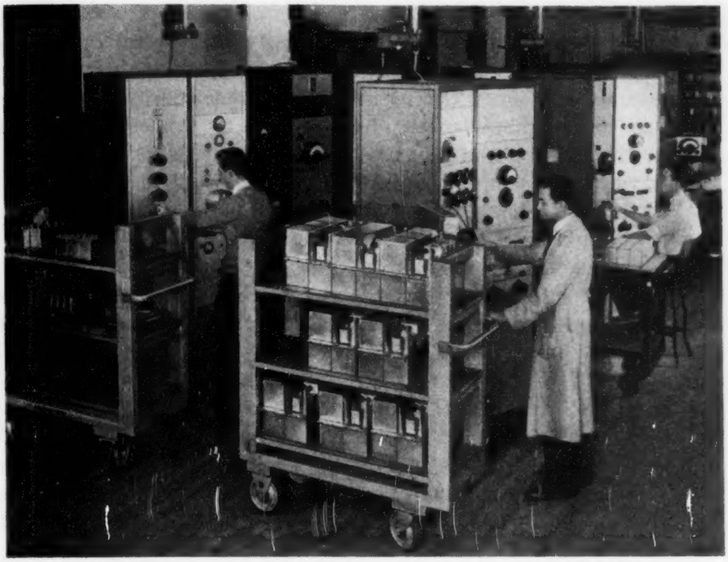


Fig. 12—Filter testing area. The standard-frequency outlets are seen above the test sets.

of each filter section is brought to a peak at a specified frequency. The filter is placed in a test shield simulating its final container. Voltage from a precision oscillator is applied to the input terminals of the section and the voltage at the output terminals is measured with a sensitive detector preceded by a variable attenuator. The condensers designated C_{TL} in Fig. 13 are adjusted with a non-metallic tool until the loss reaches a peak of 50 to 70 db. Each section contributes two such peaks.

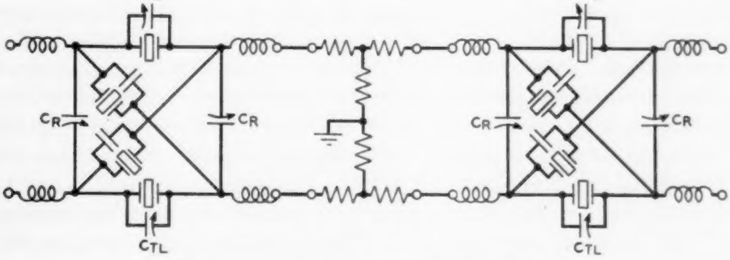


Fig. 13—Schematic of crystal channel filter.

The filter is then transferred to a resonance bridge and the impedance looking into either end of each section, with the opposite end open-circuited, is adjusted to series resonance at a specified frequency by means of the condenser C_R. This adjustment primarily controls the shape of the loss characteristic of the filter in the transmission range. The resistance at resonance is recorded for later reference.

In some types of filters, adjustment must be made to secure the correct absolute capacitance between certain points in the filter rather than to obtain desired attenuation peaks or resonances. A capacitance bridge is employed for this purpose.

In all of these adjustments, test leads connecting the filter to the test set play an important part. Shielding, balance, capacitance to ground, dielectric loss, stability and other characteristics of the leads must be carefully controlled or compensated by adjustments within the test sets in order to meet precision requirements of the order of $\pm$ 0.01 per cent.

After adjustment, in the case of the channel filter, the individual sections are connected through a resistance pad selected to complement the values of resistance measured during resonance adjustment. Uniformity of overall transmission loss, regardless of manufacturing variations in components, is secured by this means.

The completely wired filter is now placed in a copper shell and hermetically sealed with solder, except for an inlet and an outlet vent. In order to remove vestiges of moisture which might affect the crystals or other components during service life, a current of air of less than 3 per cent relative humidity is then passed through the filter for 12 hours and the vents are sealed off.

Final test consists of measurements of transmission loss at a series of frequencies in the transmission and attenuation bands of the filter, using equipment similar to that on which the peaks were adjusted. The variety of product which must be tested with these facilities demands maximum flexibility and minimum set-up time. This requirement is met with plug-in terminating impedances, pads, leads, etc., and with oscillators and detectors tuning continuously over a wide range of frequencies. Several filters of the same type are normally tested simultaneously, all being measured at one frequency before the next frequency is set up. Contact fixtures for particular types are provided when justified by quantity requirements, in order to facilitate the transfer of test leads from one filter to the next.

Transmission loss characteristics of channel filters under various conditions are shown in Fig. 14. The solid curve illustrates a normal filter. The loss in the passband is approximately 5.6 db, with distor-

tion of about 0.25 db over a band 3 KC wide. The peak losses are from 75 to 90 db, with the intervening "valleys" approximately 65 db. The other curves illustrate three types of defects occasionally observed, namely: (*a*) displacement of one attenuation peak caused by

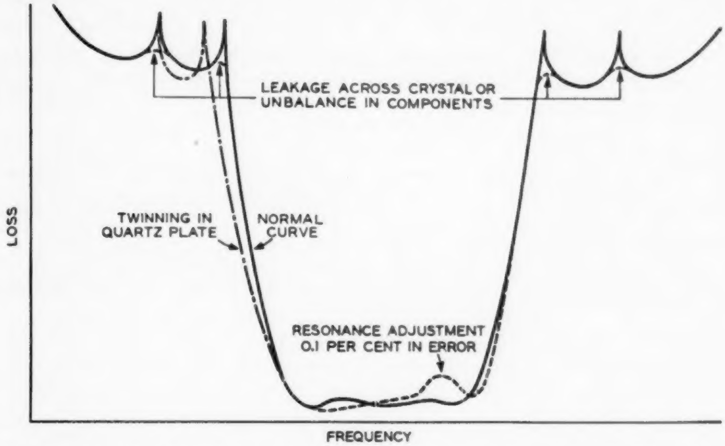


Fig. 14—Insertion loss characteristics of crystal channel filters, showing the effects of deviations from normal conditions.

twinning in one of the crystal plates, (*b*) abnormal distortion caused by a 0.1 per cent error in resonance adjustment, and (*c*) low loss at peaks caused by leakage across a crystal or by components which are inadequately balanced to ground. As an aid to locating the particular components or adjustments which are responsible for such defects, a catalogue of "trouble-shooting" instructions, arranged by classes of filters and types of symptoms, has been compiled.

Cleaning, finishing and labelling constitute the remaining operations on crystal filters. High temperature processes such as vapor degreasing and baking of the finish are inapplicable here because of the nature of the component apparatus in the filter. The surface is scratch-brushed, washed with a solvent and sprayed with aluminum lacquer. Rubber stamps and printers' ink are then used to apply the terminal and type designations.

Conclusion

Crystal filters exemplify the trend toward higher frequencies and higher precision in modern carrier systems. These advances in design have required the development of new manufacturing processes and refined methods of adjusting and testing, and demand increased care

and skill in every step. Ten years ago this technique had not reached even the laboratory stage. Today, in the commercial production of crystal filters, it has become commonplace to deal with capacitances expressed in tenths of a micromicrofarad, transmission losses in hundredths of a decibel, crystal dimensions in thousandths of a millimeter and frequency measurements in thousandths of a per cent. The attainment of such precision at moderate shop cost is the primary engineering problem in the manufacture of the higher frequency types of carrier telephone facilities.

Bibliography

1. "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, *Bell System Technical Journal*, 1934, pages 405–452.
2. "A Carrier Telephone System for Toll Cables," C. W. Green, E. I. Green, *Electrical Engineering*, 1938, pages 227–236.
3. "A Twelve-Channel Carrier Telephone System for Open-Wire Lines," B. W. Kendall, H. A. Affel, *Bell System Technical Journal*, 1939, pages 119–142.
4. "Crystal Channel Filters for the Cable Carrier System," C. E. Lane, *Electrical Engineering*, 1938, pages 245–249.

Results of the World's Fair Hearing Tests

By J. C. STEINBERG, H. C. MONTGOMERY, and M. B. GARDNER

A hearing test for musical tones formed part of the Bell System exhibit at the New York and San Francisco Fairs, and the test records obtained have made possible a study of the hearing of a large group of the United States population. The variation of hearing acuity with age and sex is described in considerable detail. Hearing is also related in lesser degree to other factors, such as place of residence, economic status, and race, and these relations are discussed.

The data are applied to the United States population by indicating certain allowances which should be made for differences between the Fair groups and the population, particularly with respect to distribution of ages and economic status.

Accuracy of the test is discussed in relation to ability of visitors to understand the test procedure, disturbing effect of background noise, and calibration of the test equipment.

Certain results of the survey are expressed in terms of ear canal pressure and equivalent free field intensity, and on this basis a comparison is made with the results of other surveys of hearing.

A criterion is given for deciding how much hearing should vary from average before being considered abnormal. Application of this criterion indicates, in the case of children, a suggestive similarity between incidence of adenoid growth as reported in medical surveys and abnormal hearing for high frequency tones.

WITH the opening of the New York and San Francisco World's Fairs in 1939, an opportunity became available for a survey of hearing in a large group of the United States population. One of the Bell System Exhibits consisted of a hearing test whereby visitors could test their hearing for tones of musical pitch. At the end of the test, each visitor was asked to permit an attendant to make a photographic copy of his hearing test card so that a study of the records might be made. Before making the copy, the attendant indicated by a check mark whether the visitor was male or female, colored or white, and to which of the five age groups, 10–19, 20–29, 30–39, 40–49, or 50–59, she judged him to belong. In all, some 550,000 photographic records were obtained, and it is estimated that about 80 per cent of the visitors who tested their hearing for musical tones cooperated in the survey. A somewhat similar test for spoken words was also provided, but the survey was concerned principally with the results of the musical tones test.

The value and usefulness of such a large collection of records is dependent very directly upon the accuracy of the test. Therefore considerable attention was given to the calibration of the hearing test equipment and to the evaluation of factors which might affect the results of the test. There seems little doubt that the records accurately portray the hearing characteristics of that section of the population taking the tests.

One of the principal objectives of the study was to determine the hearing acuity and the prevalence of defective hearing in the United States population. The visitors who tested their hearing were not a representative sample of the population with respect to factors affecting hearing. Consequently a second objective was to determine the relation of hearing to such factors as age, sex, place of residence, economic status, etc. This information is necessary in order to apply the Fair data to the whole population or to specialized groups within the population.

It is believed that the two important factors, age and sex, have been satisfactorily evaluated. Information on other factors although less complete, is sufficient to justify many applications of the data. In other applications, it is necessary to make reservations and these are described in the text.

Description of the Test

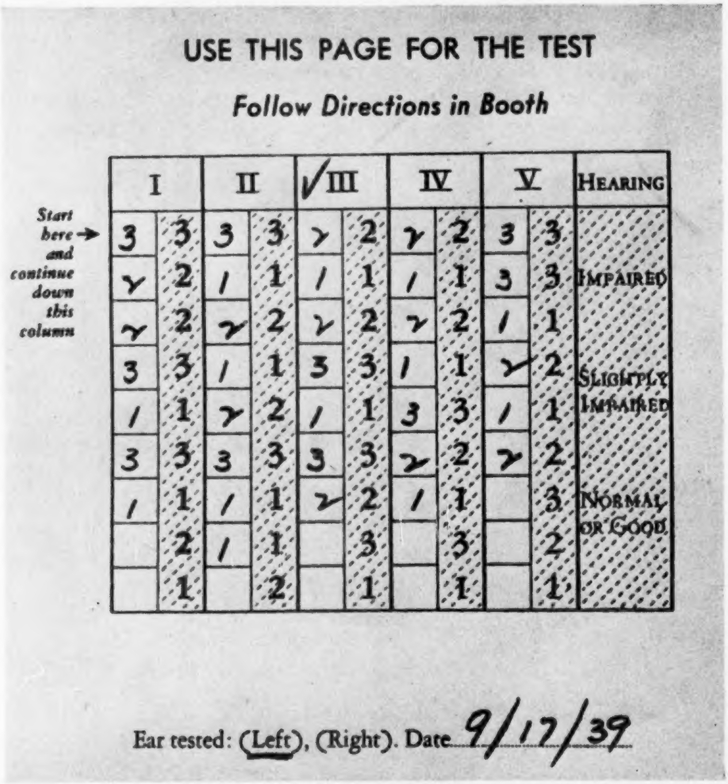
The tests were made in sound-insulated rooms arranged to seat seven visitors, each partially screened from the others, as shown in Fig. 1. The test and suitable instructions were recorded on phonograph records and given through a telephone receiver which the visitor held to his ear. In the musical tone test a pure tone was sounded one, two, or three times, and the listener was instructed to write in a space on a form that was given him the number of times he heard the tone. For a given pitch, nine such sets of tones were sounded, each set fainter than the preceding one. When the tones became too faint to be heard, the listener could not write the number correctly, and thus a measure of his hearing acuity was obtained. This test was made with tones of five different frequencies in the following order: 440, 880, 1760, 3520, and 7040 cycles. A typical hearing test record is shown in Fig. 2.

The correct numbers, which appear in the spaces between the columns of Fig. 2, were printed on the back of the blanks in such a way that they would show through in these spaces when the blank was placed on a brightly illuminated glass shelf. The designations "normal or good," "slightly impaired" and "impaired," which show through

Fig. 1—Interior view of one of the hearing test booths.

opposite the eighth, fifth, and second steps respectively, were intended to give a qualitative indication of hearing acuity. Thus the visitor could correct his own test and obtain an indication of his hearing acuity for each frequency.

A scale of hearing acuity for expressing the results of the tests was set up, using as a zero or reference level the mean test score of men and women at the Fairs in the age group 20 to 29. Hearing acuity is expressed as a hearing loss in the usual way, i.e., the departure in db of a given test result from the reference level. As shown in a later section, this reference level gives an ear canal pressure which corre-



USE THIS PAGE FOR THE TEST

Follow Directions in Booth

Start here → and continue down this column

I		II		√III		IV		V		HEARING
3	3	3	3	2	2	2	2	3	3	
2	2	1	1	1	1	1	1	3	3	IMPAIRED
2	2	2	2	2	2	2	2	1	1	
3	3	1	1	3	3	1	1	2	2	SLIGHTLY
1	1	2	2	1	1	3	3	1	1	IMPAIRED
3	3	3	3	3	3	2	2	2	2	
1	1	1	1	2	2	1	1		3	NORMAL
	2	1	1		3		3		2	OR GOOD
	1		2		1		1		1	

Ear tested: (Left), (Right). Date 9/17/39

Fig. 2—Typical hearing test record as it would appear when illuminated from underneath.

sponds closely to that for zero hearing loss on the 2A Audiometer.[1] Hence hearing losses given here are comparable in magnitude with audiometric measurements.

The range of the test is shown in Table 1, which gives the hearing loss corresponding to each test step.[2] The range covered is 62 db in the first four columns and 48 db in the last.

[1] Some confusion has been occasioned by referring to the audiometer zero as normal hearing. Actually it is supposed to represent average normal hearing, where normal hearing refers to a range about the average value. If an individual has a hearing loss, his hearing is not necessarily abnormal. The amount of hearing loss which should be considered abnormal is discussed in a later section.

[2] The hearing loss values given in Table 1 correspond to the actual tone levels in the test. Throughout this paper it is assumed that the threshold of an individual lies between the last level at which he correctly records the number of tones and the first level at which he does not, and in computing mean values the loss is reckoned mid-way between these two levels.

TABLE 1

HEARING LOSS FOR EACH STEP IN THE MUSICAL TONE TEST

Column Frequency	I 440	II 880	III 1760	IV 3520	V 7040
Step 1	52*	52	52	46	33
2	42*	42	42	36	27
3	32	32	32	26	21
4	22	22	22	16	15
5	14	14	14	8	9
6	8	8	8	2	3
7	2	2	2	−4	−3
8	−4	−4	−4	−10	−9
9	−10	−10	−10	−16	−15

* These tones were used in the instructions for the test.

The voltage levels used at each frequency in the hearing test were selected so that the average young person would be able to hear the tones on the first six or seven steps, but would miss the last two or three.

RELATION OF HEARING TO AGE AND SEX

The relation of hearing loss to age and sex may be summarized by giving the average hearing loss. More detailed information may be obtained from tables giving the frequencies of occurrence of different amounts of hearing loss. In this section, both types of data are given for men and women separately, in five age ranges.

Trends in Average Hearing

Average hearing as indicated by the mean hearing loss for men and women in five different age groups is shown in Table 2. The number

TABLE 2

MEAN HEARING LOSS IN DB

Age Group		Frequency					Number of Tests
		440	880	1760	3520	7040	
Men	10–19	1.0	.3	−.3	−1.2	−.4	4132
	20–29	.0	−.2	−.1	2.0	1.5	3287
	30–39	1.4	1.3	2.3	8.2	7.7	3197
	40–49	3.7	4.5	7.0	17.7	16.8	4528
	50–59	6.8	7.7	12.1	25.6	24.0	1935
Women	10–19	.5	.2	−1.1	−4.4	−3.6	3417
	20–29	.0	.2	.1	−2.0	−1.5	4208
	30–39	2.6	2.6	2.9	2.4	4.8	3978
	40–49	6.0	5.8	6.7	7.8	11.9	4369
	50–59	10.3	9.8	11.0	13.8	19.7	2538

of tests used in obtaining the mean values is given in the right-hand column. For each age group, they were selected in a random manner from the New York and San Francisco tests, about two-thirds from New York and one-third from San Francisco.[3]

Certain trends in average hearing are evident in Table 2. On the average, both men and women show increasing hearing impairment with increasing age. For high-frequency tones, and especially at 3520 cycles, the effect is more pronounced in men than in women, but for

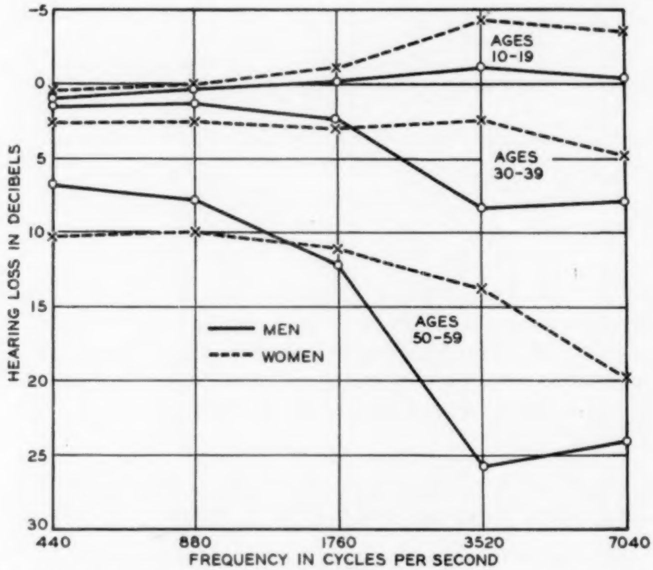


Fig. 3—Mean hearing loss in db for men and women in the youngest, middle, and oldest age groups.

low tones the opposite is true, although to a smaller degree. For the 1760-cycle tone there appears to be little difference between the hearing of men and women. These trends are shown in graphical form in Fig. 3, for the youngest, middle and oldest age groups.

At the lower frequencies, the hearing of the youngest group in Table 2 is slightly poorer than that of the next older group. It is

[3] Several sampling procedures were used, all based on selection of tests by some arbitrary rule, such as taking four tests in order then skipping twelve. In general, the same rule was used throughout a whole day's tests. The days selected were well scattered throughout the season, and week days and week-end days were used in the proper proportion. A larger sampling proportion was used for the older age groups, to make the groups in the sample more nearly equal in size.

believed that this is due principally to the greater difficulty of the younger children in understanding the test and writing their responses on the test blank.

TABLE 3

STANDARD DEVIATION OF HEARING LOSS IN DB

Age Group		Frequency				
		440	880	1760	3520	7040
Men	10–19	10.9	8.7	9.4	12.7	14.2
	20–29	8.5	7.1	8.7	13.8	14.5
	30–39	10.1	7.9	9.9	16.6	16.2
	40–49	11.6	11.0	13.4	19.3	17.0
	50–59	12.1	12.6	15.8	19.2	15.1
Women	10–19	10.6	8.6	8.6	9.4	11.6
	20–29	9.3	7.9	8.5	10.1	11.8
	30–39	11.1	9.9	10.2	12.1	13.8
	40–49	12.4	11.8	11.9	14.1	15.8
	50–59	14.0	14.0	13.9	16.2	15.4

Table 3 shows the standard deviations of the hearing losses for single tests. They range in magnitude from 7 to 20 db, and tend to increase with increasing age and tone frequency. An exception occurs for the 440-cycle tone where the values are mostly larger than for the 880-cycle tone. Since previous surveys have not shown a tendency for the standard deviation to increase below 880 cycles, it seems likely that the present increase occurred because the 440-cycle tone was the first one in the test, and initial unfamiliarity produced a greater scattering of the results. Carelessness in holding the receiver snugly against the ear would produce a similar scattering of results at this frequency.

The mean hearing losses and standard deviations for the older groups for the two high-frequency tones would be somewhat larger were it not for the restricted scale of the test. At 3520 cycles the range from zero is only 46 db, and at 7040 cycles, only 33 db. In computing means and standard deviations all results lying beyond the range of the test were grouped one test step beyond the extreme value included in the test.

Distribution of Hearing Loss

In addition to giving the trends in average hearing, the hearing test scores afford a means of determining the frequency of occurrence of different amounts of hearing loss. A convenient method of presenting the occurrence rates of different degrees of deafness is by means of curves or tables showing cumulative distributions. Table 4 shows

TABLE 4

DISTRIBUTION OF HEARING LOSS

Percentage of tests failing to show a correct response at the test step indicated, based on 23,320 tests from the N. Y. Fair and 12,269 tests from the S. F. Fair.

Frequency	Test Step	Hearing Loss	Men, Age Group					Women, Age Group				
			10–19	20–29	30–39	40–49	50–59	10–19	20–29	30–39	40–49	50–59
440	1	52*	—	—	—	—	—	—	—	—	—	—
	2	42*	—	—	—	—	—	—	—	—	—	—
	3	32	2.7	1.2	2.5	4.1	5.4	2.5	1.5	3.5	5.4	9.9
	4	22	4.6	2.1	4.2	7.1	9.8	4.1	3.0	6.1	10.7	17.9
	5	14	9.2	5.3	8.3	13.9	21.2	8.4	6.6	11.8	20.9	33.5
	6	8	16.0	11.0	14.7	22.3	32.1	15.2	12.1	19.4	30.9	44.3
	7	2	40.0	32.3	38.7	46.7	60.9	37.7	33.5	43.3	56.2	68.9
	8	−4	67.1	67.9	71.3	77.6	85.7	64.7	65.1	72.9	81.8	88.1
	9	−10	86.2	92.2	93.6	95.1	97.8	84.9	90.7	94.0	96.2	98.0
880	1	52	.4	.1	.1	.7	1.4	.4	.2	.8	1.2	1.9
	2	42	.7	.2	.4	1.8	3.2	.6	.5	1.4	2.6	5.3
	3	32	.9	.6	.9	3.2	5.4	1.0	1.1	2.1	4.1	7.6
	4	22	2.1	1.3	2.3	6.8	12.0	2.2	2.1	4.4	8.7	16.1
	5	14	4.9	2.9	5.5	13.5	21.6	4.6	4.3	8.5	16.0	27.1
	6	8	10.3	7.4	11.7	23.0	33.6	9.3	8.2	15.3	26.4	40.9
	7	2	34.3	30.2	37.5	49.3	62.3	32.5	29.9	40.8	55.1	68.0
	8	−4	77.5	76.1	81.8	87.3	92.3	77.4	78.1	84.0	89.6	93.3
	9	−10	88.9	93.8	96.2	97.6	98.9	90.6	95.0	97.4	98.2	98.8
1760	1	52	.5	.2	.4	1.7	4.3	.2	.2	.6	.8	1.6
	2	42	.6	.4	.7	2.9	6.3	.5	.4	.9	1.8	3.9
	3	32	1.0	.7	1.5	5.6	11.0	.8	.8	2.1	3.8	8.3
	4	22	1.9	1.4	3.5	11.0	19.5	1.4	1.9	4.4	8.8	16.6
	5	14	5.0	5.7	9.5	20.3	33.4	3.8	4.9	9.4	19.5	31.9
	6	8	11.3	11.5	18.3	32.9	47.8	7.9	10.7	17.8	32.9	47.5
	7	2	31.5	31.8	40.3	55.6	69.1	26.7	30.7	43.0	58.7	70.4
	8	−4	63.5	63.6	74.6	82.8	90.2	60.1	66.4	77.2	85.1	91.2
	9	−10	85.1	89.4	93.3	95.6	97.9	85.9	91.0	94.5	97.3	97.8
3520	1	46	1.8	2.5	5.8	15.3	25.6	.2	.6	1.5	3.0	6.7
	2	36	2.3	3.9	8.8	20.4	33.8	.4	1.1	2.3	5.0	10.7
	3	26	3.8	6.3	13.6	29.9	45.8	1.0	1.9	4.7	9.6	19.8
	4	16	7.8	11.7	23.5	44.8	61.7	2.9	4.5	10.4	21.9	36.2
	5	8	13.3	20.2	36.1	60.1	75.5	6.6	10.2	21.2	37.7	54.7
	6	2	26.7	35.7	54.0	73.9	86.8	16.4	21.9	38.0	56.2	72.3
	7	−4	54.5	64.1	76.9	88.8	95.3	45.0	52.5	68.6	81.7	89.7
	8	−10	78.4	85.8	93.2	97.0	99.2	71.8	81.2	91.3	95.9	98.0
	9	−16	90.6	95.6	98.1	98.9	99.7	89.0	95.5	98.2	98.9	99.3
7040	1	33	6.3	7.8	16.1	35.0	53.1	1.6	2.7	8.4	20.1	38.3
	2	27	7.1	8.9	18.1	37.5	56.6	2.0	3.3	9.6	22.7	41.1
	3	21	8.7	10.7	20.8	41.8	60.8	3.0	4.4	11.9	27.2	45.9
	4	15	13.2	15.3	28.6	50.9	70.5	6.2	7.9	19.1	38.0	57.7
	5	9	20.6	23.8	38.6	61.7	78.4	13.8	15.7	31.1	50.7	70.2
	6	3	32.5	35.7	52.4	72.6	86.4	24.1	28.2	47.2	64.9	82.1
	7	−3	49.4	54.7	70.4	85.0	94.0	41.3	49.3	67.7	81.3	92.6
	8	−9	70.2	75.6	85.0	93.8	97.5	64.0	72.3	86.0	92.7	97.6
	9	−15	82.7	88.1	94.5	97.6	98.9	80.0	87.9	95.1	97.5	99.1
No. of Tests			4132	3287	3197	4528	1935	3417	4208	3978	4369	2538

* These tones were used for the instructions for the test.

such distributions for the 35,589 test scores that were used in calculating the mean hearing loss values of Table 2. It is arranged to show, for each of the five tones, the cumulative distributions separately for men and women in the five age ranges. It gives the percentage of tests failing to show a correct response at the test step indicated, or the percentage of individuals having a greater hearing loss than that corresponding to the indicated step. For example, the table shows that only 0.7 per cent of the men in the 20–29 age group have hearing losses greater than 32 db for a 1760-cycle tone, while 11 per cent of those in the 50–59 group have this much loss.

Zero hearing loss falls between steps 7 and 8 for the first three tones, and between steps 6 and 7 for the last two tones. The last step corresponds to very good hearing, and individuals able to hear this step have hearing acuities at least 10 db better than average. Some 10 or 15 per cent of the youngest age group, but only 1 or 2 per cent of the oldest group, were able to hear the tones on the last step. For the tone of lowest frequency, there were seven young persons for every older person who could hear the last step, but at 7040 cycles, there were 18 young persons for every such older person.

The tabular data for the age groups 20–29 and 50–59 are shown graphically in Fig. 4 for four of the tones, beginning at 880 cycles.[4] The curves are cumulative distributions and the ordinate gives the percentage of individuals having hearing losses greater than the value indicated by the abscissa.[5] At 880 cycles hearing losses in excess of a given amount tend to be more prevalent among women than among men. At 1760 cycles the distribution curves for men and women are much the same. At the two higher frequencies, the prevalence of deafness in excess of a given amount is greater among men than among women.

A hearing loss of 25 db at frequencies up to 1760 cycles begins to be a handicap. The individual will usually be aware of such an impairment, and will experience difficulty in understanding speech under conditions of public address, such as in the church or theater or around the conference or dinner table. The distribution curves show that only about 1.5 per cent of the young people taking the test, or three out of 200, have a hearing loss of 25 db or more for tones of these frequencies. In the oldest age range almost ten times as many, or every seventh person, shows this much impairment.

[4] The distributions for 880 cycles may be used for 440 cycles as well, it being assumed that the small difference shown in Table 4 is due to practice.

[5] The ordinates are shown on an arithmetic probability scale, which has the property that a normal distribution plots as a straight line whose slope is proportional to the standard deviation of the distribution. It is convenient because it shows the small values more accurately and because on this scale the standard errors of the ordinates are approximately equal in all parts of the range.

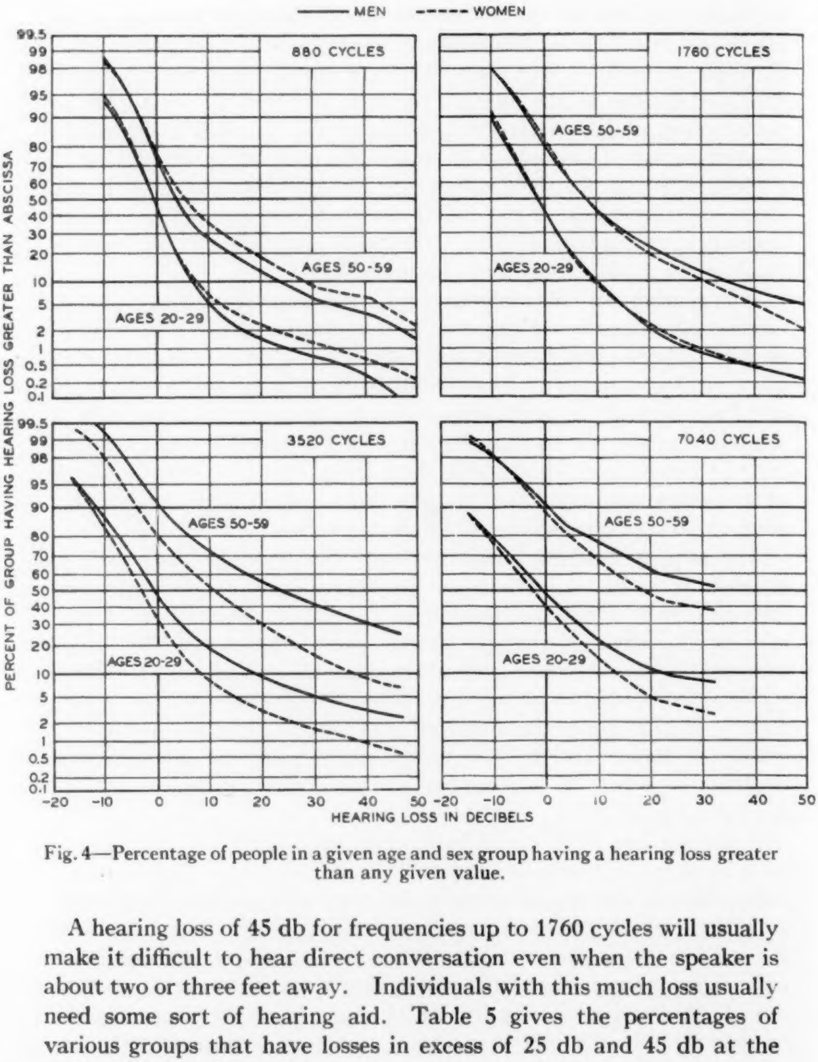


Fig. 4—Percentage of people in a given age and sex group having a hearing loss greater than any given value.

A hearing loss of 45 db for frequencies up to 1760 cycles will usually make it difficult to hear direct conversation even when the speaker is about two or three feet away. Individuals with this much loss usually need some sort of hearing aid. Table 5 gives the percentages of various groups that have losses in excess of 25 db and 45 db at the various frequencies.[4]

Acuity for the two high-frequency tones is less important than for the low tones for understanding speech, so that a loss for the high tones is not such a serious handicap. High-tone deafness is of particular interest, however, to the extent that it is indicative of a pro-

TABLE 5

PERCENTAGE OF TESTS WITH HEARING LOSS GREATER THAN 25 AND 45 DB.

Age Group		25 db Loss, Frequency				45 db Loss, Frequency		
		440, 880	1760	3520	7040	440, 880	1760	3520
10–19	Men	1.7	1.6	4.5	8.0	.6	.6	1.8
	Women	1.8	1.2	1.2	2.4	.6	.4	.3
20–29	Men	1.1	1.2	7.0	9.5	.1	.3	2.7
	Women	1.8	1.6	2.2	3.5	.4	.3	.7
30–39	Men	1.8	3.5	15.0	19.0	.3	.6	6.0
	Women	3.5	3.5	5.5	10.0	1.2	.8	1.6
40–49	Men	5.5	9.5	32.0	39.0	1.4	2.6	16.0
	Women	7.0	7.0	11.0	24.0	2.1	1.5	3.0
50–59	Men	9.5	17.0	48.0	58.0	2.6	6.0	27.0
	Women	13.0	14.0	22.0	43.0	4.0	3.0	7.0

gressive condition which may later involve tones of lower frequency. It is striking to note that, of the people taking the hearing test at the World's Fairs, some 6 per cent in the 20–29 year age range showed a hearing loss in excess of 25 db for the 7040 cycle tone. In the oldest age range half of the people showed such a loss. It is likely that an even larger proportion would be found in a random sample of the population, for, as will be discussed in a subsequent section, it is believed that the people taking the test at the Fair represent an economic status that is average or better, and there are indications of a greater prevalence of hearing defects in the lower economic groups.

The Estimation of Age

As previously indicated, estimates were relied upon to furnish information on the ages of the visitors taking the hearing test. The attendant making these estimates was changed every hour or so, about fifty being used in the course of a week at the New York Fair. In order to determine the accuracy of the estimates, 267 test blanks of members of Bell Telephone Laboratories families were examined and the real age compared with the estimated age indicated on the blank. In most cases the attendant who made the estimate was not aware that the individual was a member of the Bell System. It was found that 62 per cent of the real ages were within the estimated 10 year age group, while 83 per cent were not more than 3 years outside of the group and 96 per cent were not more than 8 years outside of the group. There was a general tendency to estimate high for young people and low for older people.

The average real age in any estimated age group depends on the distribution of ages within the group, the amount of random error involved in making the estimates, and any consistent tendency to estimate high or low. The distribution of estimated ages at the two Fairs is shown in Table 6, together with the age distribution of people

TABLE 6

PERCENTAGE OF INDIVIDUALS OVER 10 YEARS OF AGE FALLING IN VARIOUS AGE GROUPS ACCORDING TO ESTIMATES AT THE TWO FAIRS, AND FROM THE 1930 U. S. CENSUS

Age Group	New York	San Francisco	United States Population
10–19	34	27	24
20–29	25	23	21
30–39	23	23	19
40–49	14	16	15
Over 50	4	11	21
Total	100	100	100

over 10 years of age in the U. S. population (1930 census). At New York, the youngest group was considerably larger than the corresponding group in the population while the oldest group was very much smaller. The same tendency, but to a lesser degree, is seen in the San Francisco distribution. The distribution of ages fluctuated greatly on different days of the week and at different seasons. The table is based on counts on about 35 days at each Fair, well scattered through the season.

From the results obtained with the control group described above and the age distributions of Table 6, the average real age in each age group was judged to be as follows:

Age Group	10–19	20–29	30–39	40–49	50–59
Average real age	15	23	33	44	56

These values should be reliable to the extent that it is reasonable to assume that the accuracy of the age estimates of the control group was representative of the whole group at each Fair. The oldest group, which is designated 50–59 throughout this paper, included that range at New York, but at San Francisco included all persons over 50. Because of this the average real age in this group is judged to be 54 at New York and 60 at San Francisco. The value in the table is the weighted average of these two figures.

Relation of Hearing to Other Factors

In the preceding section it was shown that hearing acuity varies to quite an important extent with the age and sex of the group. The next problem is to identify any other factors to which hearing acuity is related to a significant degree. A sensitive method of determining whether such factors exist is provided by the control chart technique developed in connection with the statistical control of manufactured product.[6] When the hearing tests results, corrected for age and sex differences, were plotted on a control chart there was very definite evidence of lack of statistical control. This indicates that one or more factors exist to which hearing is significantly related, and experience in other fields in which control chart technique has been applied suggests an excellent chance of being able to identify these factors. The most straightforward procedure would have been to make a careful study of those individuals whose tests fell outside of the control chart limit and discover the factors responsible for the abnormal scores. This was not feasible at the Fairs, and other less direct methods were used as described below.

In the discussion which follows, a judgment must often be made as to whether an apparent relation between hearing and some factor under discussion is significant. The customary formula for the significance of a mean

$$\sigma_m = \frac{\sigma}{\sqrt{n}},$$

where σ_m is the standard deviation of the mean of a group of n observations and σ is the standard deviation of a single observation, has not been used, because modern statistical theory shows that this relation is valid only for data which are in a state of statistical control. The data of this section do not meet this requirement, and attempts to use the above relation as the sole test of significance are often misleading. The judgments which are expressed as to the significance of a relation are based upon the consistency with which the relation appears when the data are broken up into small groups. Space does not permit showing all the evidence on which these judgments are based, but the summaries which are presented indicate the magnitude of such relations as are judged to exist.

Place of Residence

Data from the New York and San Francisco Fairs were compared to discover any differences which might be attributed to sectional

[6] W. A. Shewhart, "Statistical Method from the Viewpoint of Quality Control" (Grad. School, U. S. Dept. of Agriculture, 1939).

TABLE 7

Difference in Mean Hearing Loss at New York and San Francisco

		Frequency					No. of Tests	
		440	880	1760	3520	7040	N. Y.	S. F.
Men	10–19	.8	.0	.0	2.4	3.6	2839	1293
	20–29	.6	−.3	−.1	1.8	3.6	2219	1068
	30–39	.8	−.4	−.4	2.6	3.1	2193	1004
	40–49	−.5	−.8	.0	4.0	2.7	3171	1357
	Aver.	.4	−.4	−.1	2.7	3.2		
Women	10–19	.0	−.8	−1.0	−.2	.5	2172	1245
	20–29	.3	−.6	−.9	−.2	1.8	2848	1360
	30–39	.7	−.5	−.9	.2	2.3	2733	1245
	40–49	.6	−.2	−.1	.4	.8	3119	1250
	Aver.	.4	−.5	−.7	.0	1.4		

differences in hearing acuity. Table 7 gives the difference in mean hearing loss between corresponding age groups at New York and San Francisco for the age groups below 50. A positive difference indicates greater hearing loss at San Francisco.

At the three lower frequencies, the differences are insignificant. The differences at the higher frequencies are not large enough to be conclusive evidence of a sectional difference in hearing, but it seems quite probable that the men taking the test at San Francisco were, on the average, some three decibels less acute than those at New York for the frequencies 3520 and 7040. No important differences in standard deviation or general form of the distribution of hearing loss were noted in comparing the two Fairs.

For approximately a week at each Fair a question was printed on each hearing test blank asking whether the visitor lived (*a*) in the city where the Fair was held, (*b*) within commuting distance, or (*c*) beyond commuting distance. From the replies to these questions it was found that roughly one-quarter of the visitors at each Fair lived in the city, and another quarter within commuting distance. The remaining one-half were probably well scattered. The test results were analyzed in relation to place of residence. Most of the differences in mean hearing loss were so small that they could easily be attributed to sampling variations. Only one of the comparisons among the various groups showed differences sufficiently large and consistent to be significant. Women from New York City had greater hearing loss at all frequencies than women from the commuting area or beyond, as shown in Table 8. The groups compared number 600 and 1400 respectively. The differences did not show any trend with

age. The table gives the average difference between corresponding age groups in the age range from 10 to 49. A similar difference was not found among the men nor among the corresponding groups at San Francisco.

TABLE 8

Frequency	440	880	1760	3520	7040
Hearing Loss Difference	1.4	2.7	2.2	1.6	2.6

The differences just discussed are small enough so that the average hearing values computed from all the data will not be critically dependent on the weight assigned to the various geographical groups. On the other hand, some of the differences are large enough to suggest that a more efficient segregation into geographical groups, taking account of past as well as present place of residence, might uncover some substantial differences in hearing.

Personal Characteristics

An attempt was made to determine the relation to hearing acuity of several such factors as economic status, intelligence, and general appearance. This was done by observing individuals at the New York Fair as they submitted their test blanks for the photographic record. Some ten or fifteen seconds of observation were usually available, and the individual was classed as below, average, or above in respect to the characteristic being studied. Although separate estimates were attempted for each of the three characteristics named above, it was concluded that in each case the same thing was being estimated, namely general personal appearance. Accordingly all the data were combined. Table 9 summarizes the findings. Each figure is the mean of the mean hearing losses for the age groups below 50.

TABLE 9

VARIATION OF MEAN HEARING LOSS WITH PERSONAL APPEARANCE

	Frequency					No. of Tests
	440	880	1760	3520	7040	
Men—Below	4.3	2.4	1.8	4.4	5.7	95
Average	1.5	1.5	1.9	8.2	7.6	560
Above	−1.2	0.5	0.8	6.0	2.0	184
Women—Below	4.4	3.2	3.8	4.3	6.7	52
Average	2.2	2.7	2.5	2.4	3.6	658
Above	0.6	1.4	1.7	0.8	2.3	259

The differences shown suggest a relation between hearing acuity and general personal appearance, although the evidence is not conclusive. In each of the ten comparisons given in the table, there is an increase in hearing acuity in going from average to above average and in nine out of ten there is an increase in going from below average to above average. Personal appearance is somewhat related to economic status and intelligence. A more accurate index of these might show a more striking relation with hearing acuity.

Race

The number of tests of negroes tabulated thus far is too small to give a satisfactory picture of the hearing trends among them. However, there is no indication of substantial departure from the results reported by Bunch and Raiford,[7] who determined that the hearing of negro men and women is similar to that of white women.

Awareness of Hearing Impairment

People whose hearing is impaired are often quite sensitive, and it seems possible that some may have avoided the hearing test for this reason. On the other hand, a person with impaired hearing might be especially attracted by the opportunity to measure it. Whether the data show too high or too low an incidence of hearing impairment depends on which of these factors predominates. This is a possibility for bias that is present in any survey where participation is voluntary. No satisfactory method of evaluating it has been discovered. However, the following discussion is intended to give some idea of the magnitude of the error which may be involved.

Since a person is scarcely aware of a hearing loss of less than 25 db, it may be assumed that neither of these factors would affect the distributions below that value. For greater losses some effect may be expected, gradually increasing so that above 40 db the possibility of a substantial bias in the distributions must be considered. The shift of the mean values of hearing loss is probably not very pronounced. For example, Table 10 shows the shift in mean hearing loss at 1760

TABLE 10

Age	20–29	30–39	40–49	50–59
Men	0.4 db	0.7 db	2.6 db	4.8 db
Women	0.4	0.9	1.6	2.9

[7] C. C. Bunch and T. S. Raiford, "Race and Sex Variations in Auditory Acuity," *Arch. of Otolaryng.*, **13**: 423–434 (1931).

cycles which results when the number of cases of hearing loss over 42 db is tripled. Eliminating all cases over 42 db would produce about half as much shift in the opposite direction. However, the form of the distribution curves for large values of hearing loss may be substantially in error.

Hearing losses at 3520 and 7040 cycles are much more common, but are not likely to be noticed except when accompanied by a loss at a lower frequency. Consequently, the biasing effect of a selective process based on awareness of hearing loss is less pronounced at these frequencies.

Right and Left Ears

The physical arrangements at the Fairs made it awkward for a right-handed person to test his right ear. As a result, about 80 per cent of the recorded tests were for the left ear. No appreciable difference was found between the test results for right and left ears. However, this may be taken as only a rough indication of equality between the ears, in view of the differences in test conditions for the two ears.

Time of Day

Data from each Fair were studied to determine whether there was any significant variation in hearing with time of day. Table 11, which gives values of mean hearing loss covering a period of about two weeks at the San Francisco Fair is typical.

TABLE 11

VARIATION OF MEAN HEARING LOSS WITH TIME OF DAY

	Frequency					No. of Tests
	440	880	1760	3520	7040	
Morning (9–1:30)	1.0	1.2	1.9	3.5	3.5	1285
Afternoon (1:30–5:30)	2.1	1.4	2.3	3.3	4.6	1637
Evening (5:30–10)	1.4	1.6	2.3	3.4	4.4	1511

The figures given are averages of several age groups. An apparent slight trend to poorer hearing in the afternoon is probably not significant, because detailed study of this and other data showed that this trend was not consistent. It was concluded that there were no trends of hearing acuity with time of day in any age group of sufficient magnitude to be revealed by the survey.

Variation Over Longer Periods

A comparison of data obtained at the New York Fair during a period early in the summer and another period early in the fall is given in Table 12. This table gives the difference in mean hearing loss for the two periods for the various age and sex groups. A positive difference indicates greater hearing loss in the later period.

TABLE 12

Difference Between Mean Hearing Losses Determined During Two Different Periods

Ages		Frequency					No. of Tests
		440	880	1760	3520	7040	
Men	10–19	.4	.2	.0	.7	.6	1509, 1330
	20–29	−.7	−.8	−.6	.9	−.2	1110, 1109
	30–39	−1.0	−.6	.5	1.1	1.4	1053, 1140
	40–49	.2	.2	.2	1.4	1.1	1802, 1369
	50–59	1.8	1.9	2.6	5.0	3.1	432, 511
Women	10–19	−1.4	.0	−.4	−.7	−.4	1028, 1144
	20–29	−1.0	.1	.5	.9	.6	1645, 1203
	30–39	−.7	.2	.3	.1	1.0	1363, 1370
	40–49	−.5	−.3	−.3	.6	1.0	1722, 1397
	50–59	.3	1.0	1.3	1.8	1.1	440, 643
Average	10–49	−.6	−.1	.0	.6	.6	

The average difference is rather small, and may be taken to indicate good stability on the part of the test equipment and lack of any pronounced seasonal trends in hearing over this interval.

Distribution of Hearing Acuity in the United States Population

The tests at the two Fairs constitute a large cross section of the United States population. It is not a representative cross section in certain respects, the most important of which are described below. It is believed that by taking into consideration the limitations mentioned below an estimate of the distribution of hearing acuity in the United States population can be obtained which is sufficiently accurate for most practical purposes.

The two Fairs taken together probably represent a good geographical cross section of the country, except that areas near the Fairs were too heavily represented. However, since no pronounced differences in hearing were found for those living near the Fairs, the geographical sampling may be regarded as fairly satisfactory. Although it is quite

possible that sections might be found in which hearing differed markedly from the Fair values, it is unlikely that such areas would be extensive enough to affect the overall result.

With regard to economic status, intelligence, and amount of education, the Fair groups were judged to be somewhat above average, and probably representative of the upper two-thirds or three-quarters of the population. If hearing is related to such factors, as seems probable, the hearing of the population is not quite so good as indicated by the Fair tests.

The portion of the distribution curves relating to large hearing losses at the three lower frequencies must be accepted with reservations on account of the possible biasing effect of awareness of hearing impairment, as discussed in the preceding section. However, the curves should be reliable below about 35 db loss, and the curves for the two highest frequencies should be reliable throughout the test range.

The distribution of ages at the Fairs was quite different from that in the population. The first step in allowing for that difference was made by recombining the distributions of hearing loss for various age and sex groups shown in Table 4, weighting each according to the size of the group in the population. The resulting distributions are shown in Fig. 5, and apply to the age range 10–59 years.[4]

In a similar manner, the figures in Table 5 for the incidence of hearing loss of 25 db or more were weighted according to the size of the groups and combined, leading to the values given in the first line of Table 13, for the age range 10–59. This process can be extended to include the whole age range as follows. It is assumed that the incidence for ages under 10 is the same as in the 10–19 group. This may not be strictly true, but is a sufficiently good approximation for this purpose. For ages above 60 a minimum estimate was obtained by assuming that the incidence of hearing loss is the same as in the 50–59 group, and a maximum estimate by assuming 100 per cent incidence above age 60. The actual value may be expected to be somewhere between these limits, which are shown in Table 13.

Except for the reservations stated in the first four paragraphs of this section, it is believed that the data of Figs. 3, 4, and 5 and Tables 2, 3, 4, 5, and 13 should apply fairly closely to the U. S. population as a whole. They may also be applied to groups in the population who are not specialized in regard to any factor related to hearing. It would be unsafe to apply them to a group of very low or unusually high economic status, college graduates, unskilled laborers, foreign groups,

[4] Loc. cit.

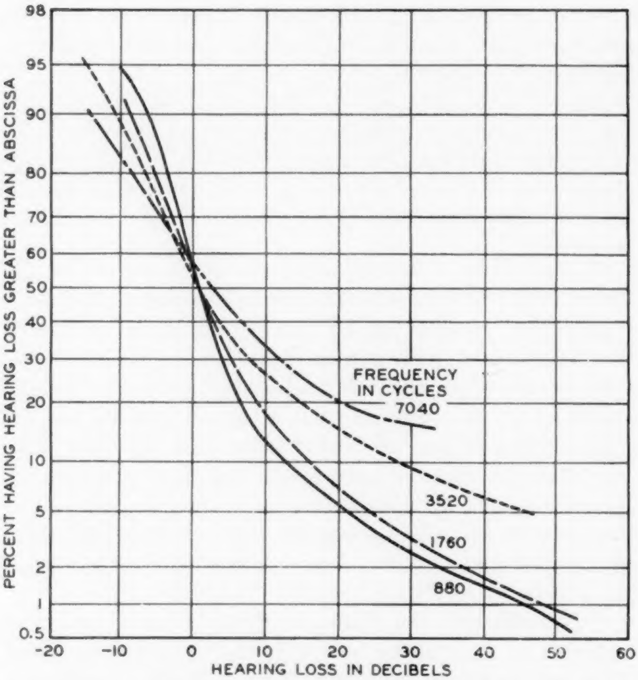


Fig. 5—Percentage of people, both men and women, in the age range 10–59 having a hearing loss greater than any given value.

TABLE 13

PERCENTAGE OF PEOPLE HAVING HEARING LOSS OF 25 DB OR MORE

		Frequency			
		440, 880	1760	3520	7040
Ages 10–59		3.8	4.5	12	18
All ages	Minimum	4	5	12	18
	Maximum	11	12	18	22

etc. without further knowledge of the relation between hearing and the factor in which the group was unusual.

ACCURACY OF THE HEARING TEST

The participation of the visitors in the test was entirely voluntary, and nothing is known from the test records concerning their reasons

for taking the test. Observation indicated that the great majority of people took the test seriously, and made a conscientious attempt to test their hearing.

It was noticed that a very small percentage of the people, mostly in the youngest age group, altered their scores by filling in all of the missing numbers before having them photographed, thus giving a false appearance of a perfect test. Some of these were detected from the differences in writing, and were eliminated from the tabulations, but others were probably included. It is believed that the number of false scores included was too small to affect the hearing loss distributions appreciably.

Some people undoubtedly secured poor scores in the test on account of failure to understand the test, interruptions, or other causes not connected with hearing. A study of this factor was made by observing about 1200 tests, picked at random, and interviewing all those who failed to fill in more than three squares of their test blanks. About 1.5 per cent failed the test in this sense, and were subsequently interviewed and watched to see if they permitted their test scores to be photographed. The interview revealed that about two-thirds of this group failed the test because they were definitely hard of hearing. Also, as it happened, about two-thirds of them submitted their test scores to be photographed, so that the number of recorded failures tended to be the same as the number that were actually hard of hearing.

The noise conditions under which the tests were given were quite favorable. The exhibit building was quiet due to the generous use of sound absorbent walls and carpeted floors. Parts of the building containing air-conditioning and other machinery were constructed on a separate foundation from the part containing the hearing test booths. The booths were carefully insulated, the attenuation of the walls to air borne sounds being 30 db or more over a wide frequency range. Additional isolation was provided by a glass partition between the booths and the lobby. Noise in the booths from external sources was nearly inaudible, and it is probable that most of the distributing noise was caused by the people participating in the test.

Sound level meter measurements with flat weighting were made in a booth where noise from external sources was judged to be most objectionable, and while regular tests were in progress. The average and maximum readings are given in Table 14. After making allowance for the attenuation of the telephone receiver covering the ear, the masking computed for the average noise level was less than 5 db at 440 cycles and zero at the higher frequencies. The masking of a

TABLE 14

SOUND LEVEL METER READINGS, WITH FLAT WEIGHTING, IN DB ABOVE .0002 DYNE PER SQUARE CENTIMETER

Frequency Band	100–300	300–500	700–900
Average reading	43	<25	$\ll 25$
Maximum reading	50	35	26

steady noise equal in magnitude to the maximum noise levels was computed to be 11 db at 440 cycles, 3 db at 880 cycles, and zero at higher frequencies.[8] The interpretation of these results is in some doubt because the people in the booths tended to be more quiet when the test level approached threshold, and also because the disturbing effect of sounds of irregular character may not be properly indicated by masking computations based on experiments with steady sounds. Accordingly a more direct method of evaluating the disturbing effect of noise was tried.

Members of Bell Telephone Laboratories who had taken the test at the Fair under routine conditions at various times during the season were retested at the Laboratories after the Fair closed. The same equipment and procedure were used, except that only one person was tested at a time under conditions free from any disturbing noise except that created by the observer himself. On the average, the tests indicated more acute hearing during the retest at the Laboratories, particularly at the low frequencies. The average shift was 2.9 db at 440 cycles, 1.4 db at 880 cycles, 1.1 db at 1760 cycles, and negligible at the higher frequencies. Since the test at the Laboratories was given last in every case these shifts may have been partly due to improvement with practice. However, they serve to set an upper limit to the average disturbing effect of noise. This comparison is based on tests of 150 ears of 106 people, whose average age was 39 and whose average hearing acuity was somewhat better than an equivalent age group at the Fair.

The equipment for the tones hearing tests consisted of eight machines at New York and two at San Francisco. These machines were maintained in an equipment room some distance from the test booths, one machine being connected to each booth. Each machine consisted of a phonograph reproducer, amplifier, attenuation network, and seven

[8] These values of masking apply to an ideal observer having a threshold approximating the minimum audible pressure curve of Fig. 6. Since a great majority of observers at the Fairs had higher thresholds, the masking would be correspondingly less for them.

telephone receivers in parallel on the output.[9] Vertical cut phonograph records contained instructions for the test and the tones used in the test. To insure a favorable ratio of signal to record and amplifier noise throughout the test, the test tones were recorded at constant level, and the desired level changes were obtained by changes in the attenuation network made in synchronism with the turntable.

Output of each phonograph reproducer and amplifier was checked daily, and held within limits which varied from ± 0.5 db at the lowest frequency to ± 2 db at the highest frequency. Performance of the attenuation networks was determined twice during the season by careful measurements of voltage at each test level and each frequency. They were found to give the expected values of attenuation within 1 db at all levels. The efficiency of each receiver was measured on a rigid closed coupler. The standard deviation of all the receivers used was about 1 db at the lower frequencies and 3 db at 7040 cycles. Check measurements were made at intervals of about one month on each receiver. The mean response of all the receivers varied by less than 1 db during the season. The ten machines were alike in output (at the test level nearest the reference level) within ± 1.0 db at the three lower frequencies and within ± 1.5 db at the two higher frequencies.

In addition to the above measurements, listening tests were made daily by one of the engineers in charge of the equipment, and the girls who conducted the test listened frequently throughout the day by means of monitoring receivers.

Hearing Test Results in Terms of Pressure and Intensity Level

In order to compare the results of the Fair tests with other data on hearing, calibrations have been made of the receivers that were used in the tests. This was done by measuring the pressure levels developed by the receiver at the opening of the ear canal for a small group of people, using a special search tube transmitter so designed that the tube could be inserted under the receiver cap into the opening of the ear canal. Such a calibration gives an ear canal pressure level in terms of receiver voltage levels. The authors are indebted to Mr. W. A. Munson of these Laboratories for the calibrations. They are preliminary in character and may need modification in the light of subsequent studies.

With the aid of these calibrations, ear canal pressure levels may be

[9] F. A. Coles, "Hearing-Test Machines at the World's Fairs," *Bell Laboratories Record*, **18**: 290, June 1940.

calculated from the receiver voltage levels measured in the tests. Such calculations for the reference level or condition of zero hearing loss as used in this paper are shown in Table 15. The resulting reference ear canal pressure levels are plotted in Fig. 6. For comparison,

TABLE 15

CALIBRATION OF HEARING TEST EQUIPMENT AT THE REFERENCE LEVEL

	Frequency				
	440	880	1760	3520	7040
Reference voltage level across receivers—db above one volt	−104	−112	−115	−114	−76
705A receiver calibration—db above 0.0002 dyne per sq. cm. per volt	133	134	133	134	98
Reference ear canal pressure level—db above 0.0002 dyne per sq. cm.	29	22	18	20	22

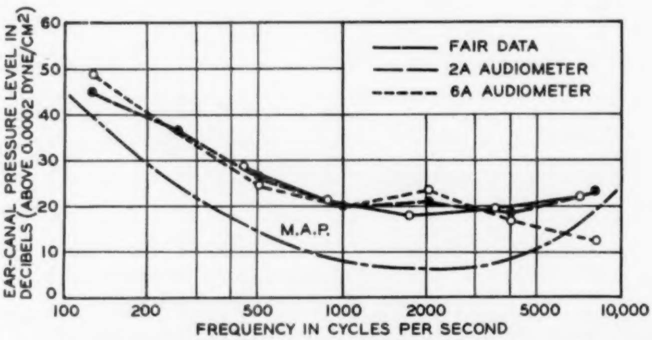


Fig. 6—Ear canal pressure level for certain reference conditions. The measurements for the M. A. P. curve were made nearer the ear drum than those for the other curves. See text.

the ear canal pressure levels corresponding to zero hearing loss on the 2A and 6A audiometers [10] and the minimum audible pressure curve derived by Sivian and White [11] are shown. The audiometer

[10] J. C. Steinberg and M. B. Gardner, "Auditory Significance of Hearing Loss," *Jour. Acous. Soc. Amer.*, **11**: 270 (1940).

In using the Audiometer it is customary to record as the hearing loss the lowest dial setting at which the tone is heard. Threshold would, on the average, be half a dial step lower than the recorded setting. Hence the curves given here for zero hearing loss are 2.5 db lower than those given for zero dial setting in the reference. See also footnote 2.

[11] L. J. Sivian and S. D. White, "Minimum Audible Sound Fields," *Jour. Acous. Soc. Amer.*, **4**: 288–321 (1933).

curves and the Fair curve are based on ear canal pressures measured at the ear opening in the manner just described. The minimum audible pressure curve is based on pressures measured about 1 cm. from the ear drum, which correspond more nearly to ear drum pressure levels. The two types of measurements are undoubtedly quite comparable below 1000 cycles. For frequencies above 5000 cycles and possibly around 2000 cycles it is believed that the pressure at the ear opening is somewhat smaller than the corresponding ear drum pressure.

A comparison of the Fair data with data from two other surveys of hearing is shown in Fig. 7. One curve shows the mean threshold

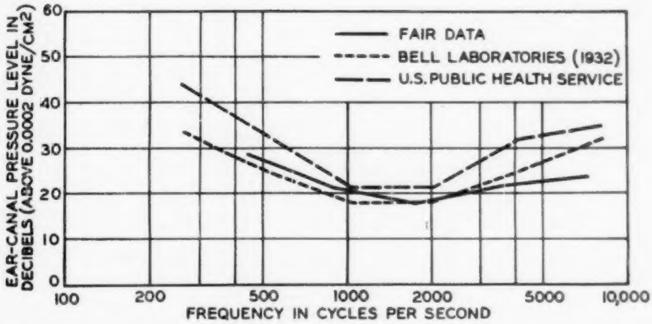


Fig. 7—Comparison of several surveys of hearing, giving mean ear canal pressure level for men aged 20–29.

pressure for men in the 20–29 age group from the Fair data. Another gives values for the same age and sex group in a survey conducted in 1936 by the United States Public Health Service using the 2A Audiometer.[12] This curve is for a somewhat selected group, including only individuals who stated when the test was made that they believed their hearing was normal. The third curve is for members of Bell Telephone Laboratories in the same age and sex group, who were tested in 1931 with a 2A Audiometer.[13] In comparing these results, it should be remembered that differences may be due to three general causes. The groups of people tested may have differed in hearing acuity. The calibrations by which the ear canal pressures were established are subject to error, especially at high frequencies. The conditions of the test, including technique, concentration of subjects, receiver fit, and background noise, were not alike in all cases. Con-

[12] W. C. Beasley, *National Health Survey, Hearing Study Series, Bulletin 5, Table 3*, The United States Public Health Service, Wash. D. C. (1938).

[13] H. C. Montgomery, "Do Our Ears Grow Old," *Bell Laboratories Record*, **10**: 311 (1932). Note that median values were given in this reference, differing slightly from the mean values used here.

sidering all the possible causes of variation the curves seem to be in fairly good agreement.

In order to give a preliminary picture of the prevalence of deafness in terms of free field intensity and frequency, the distribution curves of Fig. 5 were converted into the contour lines shown in Fig. 8. For

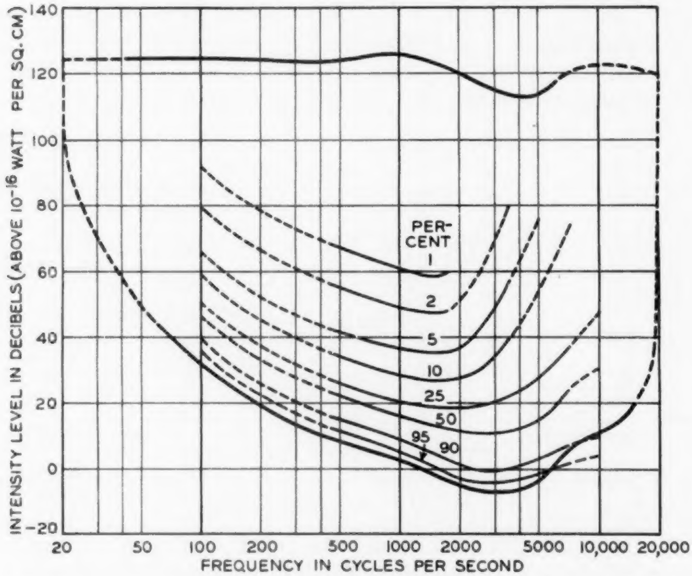


Fig. 8—Contour lines above which lie a given percentage of thresholds for the age group 10–59.

the low frequencies, this conversion was made by applying the differences between the minimum audible pressure and minimum audible field curves of Sivian and White [11] to the ear canal pressure levels of Table 15. For the high frequencies, the conversion was based on a free field calibration of the receivers. The free field intensity levels apply for the condition in which the observer faces the source and listens with both ears. The resulting contours purport to show the percentage of people in the population within the age range 10 to 59 years who cannot hear tones below the given level. The boundary lines forming the auditory sensation area represent the picture of the limits of useful hearing, based upon earlier studies.[14]

The solid portions of the contour lines represent the distributions

[14] H. Fletcher, "Auditory Patterns," *Reviews of Modern Physics*, **12**: 47–65 (1940).

obtained from the results of the Fair tests. The dotted portions represent extrapolations of the distributions beyond the intensity and frequency ranges used in the tests, and are of course speculative in character. The extrapolations to lower frequencies are based on the shape of the contours below 1760 cycles and the high correlation that has been found to exist in individual audiograms for frequencies from 64 to 512 cycles.[15,16] The extrapolations to large hearing losses for the 3520- and 7040-cycle tones were made by extending the curves of Fig. 5 as suggested by comparison with the results of other surveys.

The contours show several interesting things. The range over which hearing acuity varies is quite uniform up to about 2000 cycles, and 90 per cent of the group lie within a range of 30 db. Above 2000 cycles the range increases rapidly. Since most of the sounds met with in daily life have intensity levels greater than the 25 per cent contour, fully three-fourths of the people can hear ordinary sounds throughout the frequency range from 100 to 10,000 cycles.

The Onset of Deafness

With the increasing attention being given to the prevention of deafness by early detection, it is of considerable practical importance to define the beginning, or onset, of deafness. Perfectly normal ears are not exactly alike; some are more acute and others are less acute than the average. How much less acute than average may an ear become before deafness begins? In a preceding section, hearing loss was evaluated on the basis of the handicap that it would impose. In this section we take a different viewpoint and use the term "beginning of deafness" to mean a departure from average hearing acuity sufficient to justify the expectation of associating the departure with a specific cause.

In Fig. 9 there is shown a typical distribution curve for hearing loss. It shows the relative frequency of occurrence of various degrees of hearing acuity among a given class of people. Such curves can be well described for many practical purposes by two quantities, the average hearing loss and the standard deviation. The latter quantity, designated as σ, is a measure of the spread of the individual values from the average.

The experience of statisticians with distributions of observations of widely different character indicates that there is little chance of assigning specific causes for the deviations of observations which lie closer

[15] E. G. Witting and W. Hughson, "Inherent Accuracy of Repeated Clinical Audiograms," *Laryngoscope*, **50**: 259 (1940).

[16] W. C. Beasley, "Correlation Between Hearing Loss Measurements," *Jour. Acous. Soc. Amer.*, **12**: 104–113 (1940).

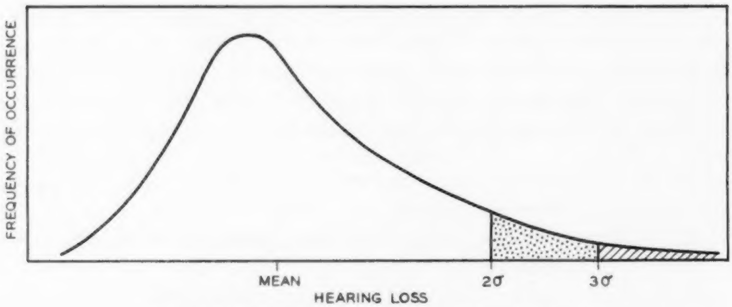


Fig. 9—Typical distribution curve of hearing loss.

than 2σ to the average, because, in this range, many causes are operating and none is predominant. In the range from 2σ to 3σ, the dotted area of Fig. 9, there is a good chance of isolating causes, although the attempt might not be justified if it involved great discomfort, danger or expense. Beyond 3σ, the cross-hatched area of Fig. 9, there is an excellent chance that the cause, or causes, can be isolated.[17] This does not imply that findable causes of hearing impairment do not exist in individuals lying closer to the average than these limits, but merely that the hearing test is not useful in selecting them. The identification of such causes and their treatment are, of course, medical problems.

TABLE 16

Hearing Loss at Which Deafness Begins, in the Sense Given in the Text

Limit		Frequency			
		440, 880	1760	3520	7040
2σ	Boys	14	15	19	22
	Girls	14	13	11	15
3σ	Boys	21	22	29	34
	Girls	21	20	18	24

This criterion leads to the limits shown in Table 16, beyond which hearing loss is significant in the sense described. Limits are given

[17] It is important to note that the standard deviation to be used in fixing the limits is to be determined after eliminating the effects of impaired hearing. Methods by which this can be done have been developed in connection with statistical methods of manufacturing control. See, for example, W. A. Shewhart, "Economic Control of Quality of Manufactured Product" (Van Nostrand, New York, 1931). The values of σ used in fixing limits in this section are 20 per cent smaller than those given in Table 3.

only for the youngest group because we are primarily concerned with early detection of deafness. Limits for older groups can be obtained in a similar manner.

The table indicates that a smaller hearing loss is significant at low than at high frequencies. At the higher frequencies a smaller loss is significant for girls than for boys. The percentage of the youngest group at the Fairs falling outside of the 2σ limits is given in Table 17.

TABLE 17

PERCENTAGE OF CHILDREN WITH A SIGNIFICANT AMOUNT OF HEARING LOSS

	Frequency			
	440, 880	1760	3520	7040
Boys	4	5	7	8
Girls	5	5	5	6

Use of the 3σ limits would lead to percentages about half as great. It is of interest to note that the 2σ limits are smaller than the amounts of hearing loss which ordinarily constitute a handicap. In other words, a hearing acuity test may have diagnostic significance before the hearing loss is great enough to produce appreciable functional impairment.

Incidence of High-Tone Deafness and of Adenoid Growth

The characteristic difference between men and women in the hearing of high tones seems now to be well established. It has been previously observed by Bunch and Raiford,[7] by Ciocco [18] and more recently by Beasley in a survey of hearing during the health census of 1936.[19] The reason for the difference is not known. It may be occupational in part, although the difference is in evidence in the youngest age group. Recent work reported from Johns Hopkins University,[20] indicates that children with high-tone deafness frequently show pronounced lymphoid tissue growth at the openings of the Eustachian tubes in the throat. In a number of cases, the deafness was cured or held in check by removing the tissue and inhibiting its growth by irradiation with radium during the adolescent period.

If there is a connection between adenoid growth and high-tone deafness, one would expect a greater incidence of adenoid growth in

[18] A. Ciocco, "Observations on the Hearing of 1980 Individuals," *Laryngoscope*, **42**: 837–856, Nov. 1932.

[19] W. C. Beasley, *National Health Survey, Hearing Study Series, Bulletins 5 and 6*, United States Public Health Service, Wash. D. C. (1938).

[20] S. J. Crowe and J. W. Baylor, "Prevention of Deafness," *Jour. Amer. Med. Assn.*, **112**: 585–590, Feb. 1939.

boys than in girls. The results of physical examinations of school children conducted by public health officers and doctors in different parts of the country indicate that such is the case. In all but one of seven surveys.[21] involving more than 18,000 children, the occurrence of adenoid growth was more frequent in boys than in girls. On the average, 6 per cent of boys and girls from 4 to 18 years of age showed pronounced adenoid growth. The ratio of the percentage of girls to the percentage of boys having the defect had an average value of 0.68; for every 100 boys affected there were only 68 girls similarly affected. Analysis of the World's Fair hearing test records shows that 7 per cent of boys and girls from 10 to 19 years of age are deafened for a 7040-cycle tone to the extent of the 2σ limit described in the previous section. The ratio of the percentage of deafened girls to the percentage of deafened boys is 0.75, or 75 girls for every 100 boys. Although it should not be concluded that the similarity of these ratios establishes a correlation between adenoid growth and high-tone deafness, it is believed that they are sufficiently suggestive to justify further study of these defects.

Acknowledgment

This survey was made possible by the cooperation of a large number of people in many parts of the Bell System. The planning, design, and construction of the exhibit were shared by the American Telephone and Telegraph Company, The Western Electric Company, Electrical Research Products, Inc., and Bell Telephone Laboratories, Inc., and to them the authors are indebted. We wish to express our gratitude to the Pacific Telephone and Telegraph Company and the New York Telephone Company for their efficient operation of the exhibits and for the large share which they had in obtaining the data used in the survey; to the tabulating and mathematical groups at the Laboratories for many hours of painstaking labor in treating the data; and to many of our associates whose suggestions and criticisms were a valuable aid in the analysis of the information. We also wish to express our appreciation to the large group of interested visitors to the Fairs whose participation in the hearing tests constituted the basic material of this survey.

[21] "The Health of the School Child," *Public Health Bulletin 200, Table 46, page 141,* United States Public Health Service, Wash., D. C. W. Franklin Chappel, "Examination of the Throat and Nose of 2000 Children," *Jour. of Med. Sciences*, **97**: 148–154 (1889). Wm. R. P. Emerson, "Physical Defects in 1000 Children," *Amer. Jour. of Diseases of Children*, **33**: 771–778 (1927).

The Subjective Sharpness of Simulated Television Images

By MILLARD W. BALDWIN, JR.

1. Introduction and Summary

OF the many factors which influence the quality of a television image, the one which is generally indicative of the value of the image and the cost of its transmission is the resolution, or sharpness. This resolution factor has always been reckoned in purely objective terms, such as the number of scanning lines, or the number of elemental areas in the image, or the width of the frequency band required for electrical transmission at a given rate. The subjective value of sharpness has not previously been considered. Some recent tests with a small group of observers, using out-of-focus motion pictures in a basic study of the visual requirements on images of limited resolution, have thrown new light on the evaluation of resolution and sharpness. The results appear of sufficient interest, particularly when interpreted in terms of television images, to warrant this presentation. We shall use the word *sharpness* in the sense of a subjective or psychological variable, with a strict technical significance in keeping with our experimental method, and we shall use the word *resolution* in the sense of an objective or physical variable.

We find that as images become sharper, their sharpness increases more and more slowly with respect to the objective factors. We find also that as images become sharper the need for equal resolution in all directions becomes less and less, and that with images of present television grade the tolerance for unequal horizontal and vertical resolutions is already remarkably wide. These conclusions are supported by our experiments with small-sized motion pictures viewed at a distance of 30 inches, about 4 times the picture height. It would not be safe to extrapolate the results of these experiments to the large-screen conditions of motion-picture theaters, because the visual acuity of the eye may be expected to increase with distance in the range in question,[1] and for other reasons.

2. Exposition of Method

Image sharpness is to be measured by subjective test, employing psychometric methods [2] which have been widely used in the measurement of other subjective values. Test images are to be projected onto a screen from 35 mm. motion picture film in such a way that the reso-

lution of the image can readily be varied over a substantial range, and with provision for making the horizontal resolution different from the vertical. The use of motion pictures instead of actual television images permits sharpness to be studied independently of other factors, and facilitates the experimental procedure.

The relationship between the television image and the motion picture which simulates it will be determined on the basis of their subjective equality in sharpness. For that purpose, a television image reproduced by an apparatus * of known characteristics is to be compared with a projected out-of-focus motion picture of the same scene, under the same conditions of size, viewing distance, brightness and color. (The motion picture will in general be superior in the rendition of tone values and in respect to flicker, and will of course not show the scanning line structure of the television image or any of the degradations commonly encountered in electrical transmission.) When the two images are judged to be equally sharp by the median one of a group of observers, the size of the figure of confusion of the motion picture is to be taken as the measure of the resolution of the compared television image.

The figure of confusion of the motion picture is that small area of the projected image over which the light from any point in the film is spread. Every point produces its own figure of confusion, of proportionate brightness, and the overlapping of the figures in every direction accounts for the loss of sharpness. When the projection lens is "in focus," the figure of confusion is a minimum one set by the aberrations of the optical system and by diffraction effects. As the lens is moved away from the "in focus" position, the figure of confusion becomes larger and assumes the shape of the aperture stop of the projection lens. If the illumination of the aperture stop is uniform, this larger figure of confusion is a well-defined area of uniform brightness. We used a rectangular aperture stop, at the projection lens, whose height and width could be varied reciprocally so as to maintain constant area of opening, and we used a calibrated microscope to measure the departure of the lens from the "in focus" position. Thus we could produce images of various degrees of sharpness and of unequal horizontal and vertical resolutions.

This method of specifying the resolution of an image in terms of the size of the figure of confusion affords an important advantage. It avoids the necessity for postulating any particular relation between the resolution and the spatial distribution of brightness values about

* The television apparatus comprised a mechanical film scanner and an electronic reproducing tube designed specifically for television. A description of it is given in reference 9.

originally abrupt edges in the image. The variety of such relations assumed by others[3,4,5,6,7] has led to a variety of conclusions with respect to resolution in television. We find subjective comparison of images to yield results of fairly small dispersion.

Let us consider now the measurement of sharpness in subjective terms. Here we find no familiar units of measurement, no scales or meters. We find no agreement as to the meaning of a statement that one image looks twice as sharp as another. We can say of two images only that (*a*) one image looks sharper than the other, or (*b*) the two images look equally sharp. When the images are quite different, there will be agreement by a number of observers that the one image is the sharper. When the images are not different in sharpness, there may be some judgments that one of them is the sharper, but these will be counterbalanced in the long run by an equal number of judgments that the other is the sharper. When the images are only slightly different in sharpness, an observer may reverse his judgment from time to time on repeated trials, and he may sometimes disagree with the judgment of another observer. It is within this region of small sharpness differences, in the interval of uncertain judgments where the observer is sometimes right and sometimes wrong with respect to the known objective difference, that it becomes possible to set up, on a statistical basis, a significant quantitative measure of sharpness difference.

Suppose that in judging two images of almost equal sharpness the observers have been instructed to designate either one or the other of the images as the sharper; that is, a judgment of "equally sharp" is not to be permitted for the present. An observer who discerns no difference in sharpness is thus compelled to guess which image is sharper, and his guess is as likely to be right as it is to be wrong, with respect to the known objective difference. Suppose, further, that the sharpness difference has been made so small that only 75 per cent of the judgments turn out to be right, the remaining 25 per cent being wrong. On the basis that these wrong judgments are guesses, we must pair them off with an equal number of the right judgments, so that 50 per cent of the total are classed as guesses. The other 50 per cent are classed as real discriminations. (The pairing of an equal number of the right judgments with the wrong judgments goes back to the equal likelihood of right and wrong guesses; it affords the best estimate we can make of the number of guesses.) When real discrimination is thus evidenced in one half of the observations, that is, when 75 per cent of the judgments are right and 25 per cent of them are wrong, we shall designate the difference in resolution as the *difference limen.**

* The term *limen* is frequently used in psychometry in lieu of older terms such as *just-noticeable-difference, threshold value, perceptible difference,* etc. It has the virtue

It is seen that the value of the limen is arrived at statistically, taking into account the variability of individual judgments. Smaller differences than the limen are not always imperceptible, nor are larger differences always perceptible.

The difference in sharpness, or in sensory response, which corresponds to a difference of one limen in resolution may be said to be one unit on the subjective scale of measurement. We shall designate this as a *liminal unit.** It will be understood that the word *liminal* has here a particular and precise significance, by reason of the one-to-one correspondence between the liminal unit and the statistically-derived value of the difference limen. A liminal unit of sharpness difference may be considered as the median of a number of values of sensory response to a difference of one limen in resolution.

3. Sharpness and Resolution

Figure 1 shows how we find the sharpness of an image to vary as the number of elemental areas in the image is changed. Sharpness is expressed in liminal units, based on measurements of the limen at four different values of resolution, indicated by the four pairs of points on the curve. Resolution is expressed as the number of figures of confusion in a rectangular field of view whose width is 4/3 of its height and which is viewed at a distance of 4 times its height.† This conventional field of view was chosen as typical of viewing conditions for motion pictures and television images. (The conventional field is 19° wide by 14° high.) The range of the curve in Fig. 1 may be stated very roughly as from 150-line to 600-line television images.

The significant feature of this curve is its rapidly decreasing slope with increasing sharpness. It shows that sharpness is by no means proportional to the number of elemental areas in the image, and demonstrates that the use of objective factors as indices of sharpness should be regarded with more than the usual amount of caution. It shows

that its meaning may be precisely defined in terms of the particular experimental method under consideration, without the extraneous significance which might attach to the more commonplace words.

* There seems to be no accepted name for such a unit. Guilford[2] calls it simply "a unit of measurement on the psychological scale." In discussing the measurement of sensory differences which are equal to each other but not necessarily of liminal size, the terms "sensory value" and "scale value" have been used.

† We have used relative values here in order that our results might be applied to other images not too different in size from the small ones we actually used. Other values of aspect ratio in the neighborhood of 4 to 3, and other values of viewing distance in the neighborhood of 4 times the picture height, may be brought within the scope of our data on the assumption that the sharpness is the same if the solid angle subtended by the area of the figure of confusion is the same. For example, a square field of view containing 60 thousand figures of confusion, and viewed at 5 times its height, would be equal in sharpness to our conventional field containing 125 thousand figures of confusion [$125 = 60 \times 4/3 \times (5/4)^2$].

that images of present television grade are well within a region of diminishing return with respect to resolution, a region, however, whose ultimate boundary is still well removed. (We estimate that the sharpest image our motion picture machine could project would be repre-

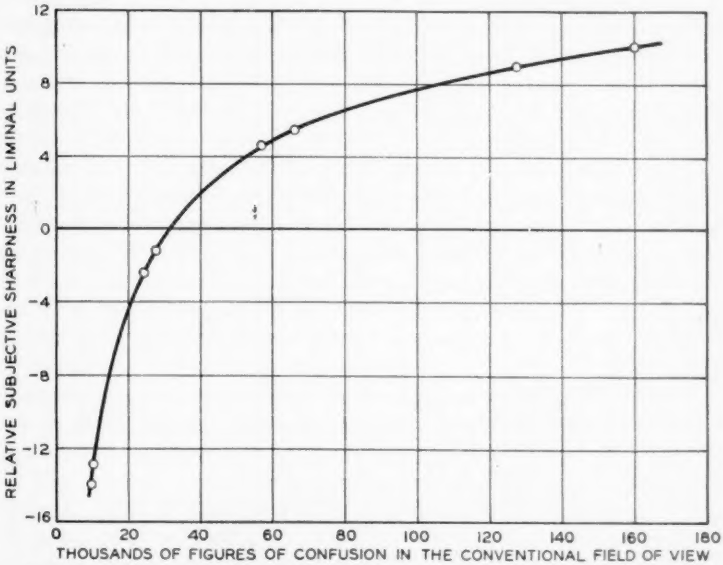


Fig. 1—Sharpness of small-sized motion pictures as a function of resolution. The conventional field of view is a rectangle whose height is 1/4 the viewing distance and whose width is 4/3 the height. Reference sharpness is approximately that of a 240-line, 24 frame per second, 806 kilocycle television image. Curve based on 1,080 observations at a viewing distance of 30 inches.

sented in Fig. 1 by a point in the neighborhood of + 18 units.) It must be remembered that the curve represents judgments made by trained observers under optimum conditions for distinguishing small differences, and that a change as small as one liminal unit, under the conditions of ordinary television viewing, would probably be largely unnoticed.

A better understanding of the meaning of this curve relating sharpness to resolution may be had by examining the experiment in detail. An individual observation was made when one of the observers, watching the projected image, caused the projection lens to be moved from a reference position to a neighboring one and reported which position he judged to yield the sharper image. The motion picture scene was a close-up of a fashion model turning slowly against a plain neutral

background, and was repeated every quarter minute. The observer could have the lens moved whenever and as often as he wanted to before reporting, so that he soon acquired the habit of observing only the most critical portions of the scene. As soon as his report was recorded, completing that observation, he was shown a new pair of lens positions, the same reference one with a different neighboring one, and asked again to report which he judged to yield the sharper image.

We believe that there were no contaminating influences and that only the size of the figure of confusion was varied. No change in brightness or in magnification could be detected. A minute lateral shifting of the image, because of play in the focusing mount of the lens, was completely masked by the continual weave of the film in the gate and the natural motion of the model. Any significance of the position of the observer's control key was destroyed by reversing its connections from time to time, between observations, without the observer's knowledge. No tell-tale sound accompanied the small motion of the lens, and none of the operator's movements could be seen by the observer.

Each one of 15 observers made 84 separate observations of sharpness difference. Expressing the resolution in terms of the angle at the observer's eye subtended by the side of the square figure of confusion, there were four main reference values, namely 0.71, 1.1, 1.7 and 2.8 milliradians (1 milliradian is equal to 3.44 minutes of arc). At each of these reference values there were seven neighboring values, namely 0, 0.045, 0.090, 0.13, 0.18, 0.22 and 0.27 milliradians greater than the reference value. (The 0 in that set means that the reference value was shown against itself, or that the observer was asked to judge a null change; this was intended to keep him on his guard and alert, not to furnish primary data.) Each pair of values was presented to each observer three times, so that there were 45 observations on every pair. The pairs were presented in irregular order according to a schedule, the variation about one reference value being completed before going on to the next. The differences were set up on the basis of preliminary trials to include some which almost none of the observers could detect and some which almost all could. It was explained that some of the differences to be judged would probably be too small for discernment, and that a "no choice" response would be permitted whenever reasonable effort failed to establish a definite choice.

The primary data are shown in Fig. 2. Each point shows the proportion of the observations in which the variable image, which had the poorer resolution by reason of its larger figure of confusion, was nevertheless judged to be sharper than the reference image. Such a judg-

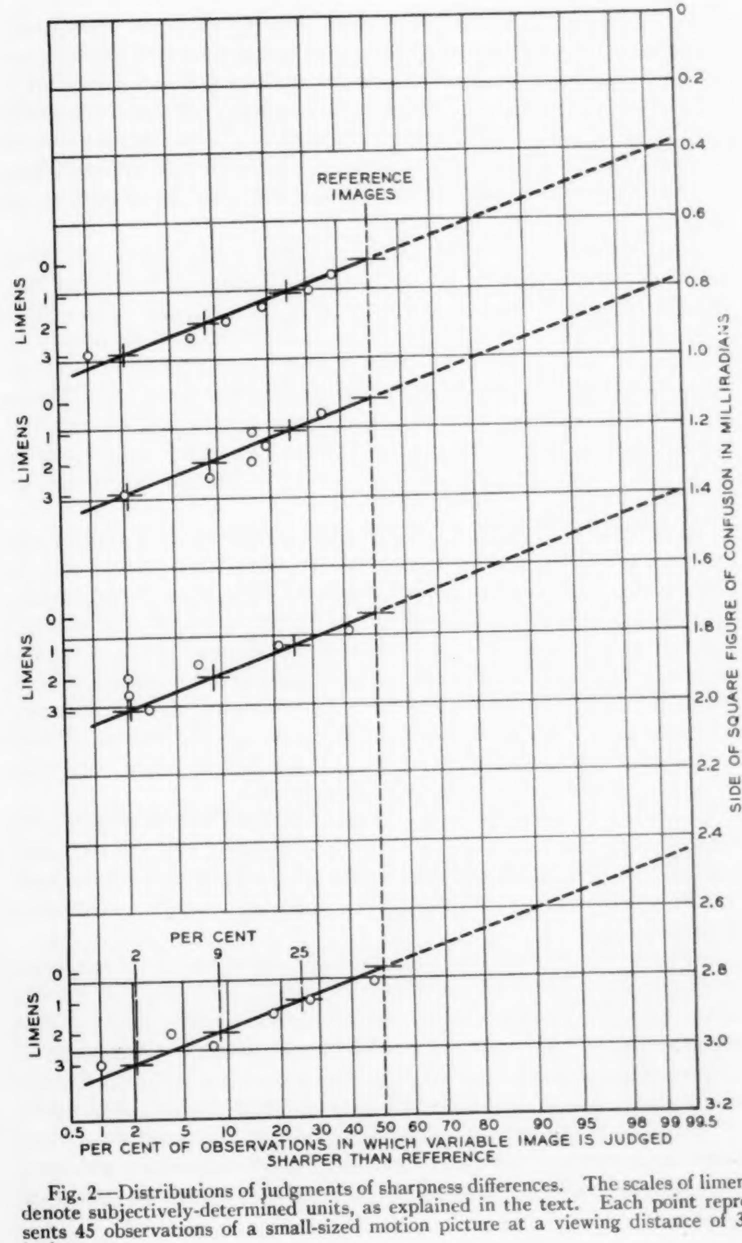


Fig. 2—Distributions of judgments of sharpness differences. The scales of limens denote subjectively-determined units, as explained in the text. Each point represents 45 observations of a small-sized motion picture at a viewing distance of 30 inches.

ment would, of course, be classed as wrong. All reported "no choice" judgments have been distributed equally between the "right" and "wrong" classes. It will be noticed that there was some discrimination at even the smallest change made, that is 0.045 milliradian, and that there was lack of complete discrimination at the largest change, that is 0.27 milliradian. The "no choice" judgments comprised 15 per cent of the total at the smallest change and only 2 per cent at the largest.

It is interesting to note that for the null changes the "no choice" judgments comprised only 17 per cent of the total, indicating either that the observers were reluctant to admit that they were guessing or that they were judging coincidental small changes in the film due to its bending in the gate or to its photographic processing. (We did observe, in establishing the lens position for sharpest focus, that film at the start of a reel required a slightly different lens setting from that at the end of the reel, and we ascribed it to the varying tension during projection, or to the varying degrees of curvature in storage on the reel.)

The four sets of points in Fig. 2 exhibit rather striking similarities. Each set may be fairly represented by a normal error curve (straight line on this arithmetic probability paper). We have drawn in four such normal curves, passing each one through the 50 per cent point at the null change and giving each a common slope. The appropriate value of slope was determined by inspection of an auxiliary plot in which the four reference values were superimposed and the four sets of points were plotted to a common ordinate scale of differences. These normal curves are considered to represent the data as well as any more elaborate relations that might have been used.

We varied the resolution only in the direction of decreasing it with respect to the reference values. We presume that had the variation been in the opposite direction the data would have been represented equally well by the same normal curves, which are accordingly extended in dotted lines.

In Fig. 2 we have indicated the magnitude of a difference of one limen by means of supplementary scales of ordinates. Since the four normal curves have a common slope, the difference limen turns out to have a constant value, 0.090 milliradian (0.3 minute of arc), independent of the size of the figure of confusion in the range from 0.71 to 2.8 milliradians. Why this should be so is a problem of physiological optics which is rather beyond the scope of this paper. The supplementary scales also serve to illustrate the meaning of differences two and three times as large as the difference limen. That is, a change in the side of

the figure of confusion of 0.18 milliradian would be twice as large as the limen of 0.090 milliradian, and would result in wrong judgments in 9 per cent of the observations, corresponding to real discrimination in 82 per cent of them. Likewise a change of 0.27 milliradian would result in real discrimination in 96 per cent of the observations. Any change larger than about three times the limen would be discriminated in practically every instance, under the conditions of our experiment.

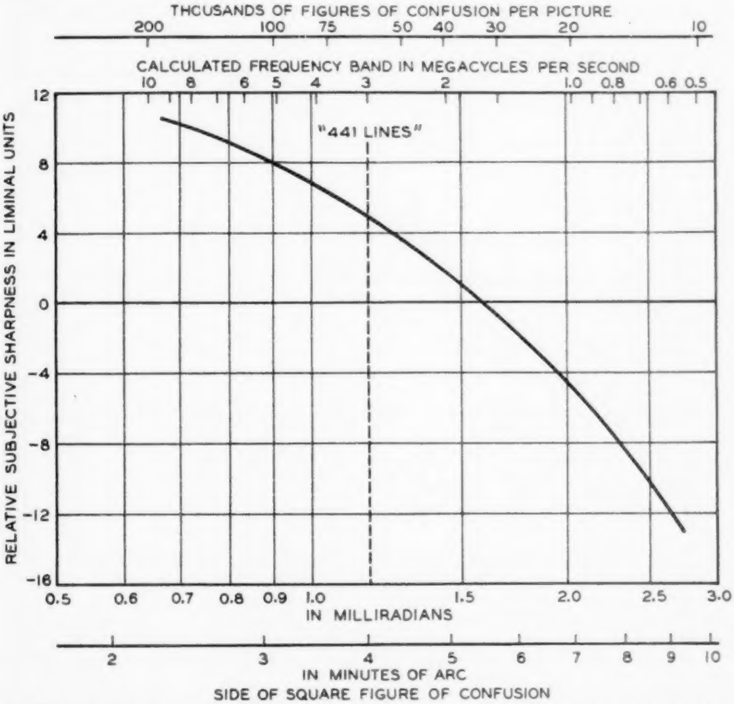


Fig. 3—Sharpness of small-sized motion pictures at a viewing distance of 30 inches. The frequency band is calculated on the basis of a 10-inch by 7½-inch television picture, 30 frames per second, with 15 per cent horizontal and 7 per cent vertical blanking, under the condition of equal horizontal and vertical resolutions.

Figures 3 and 4 show the curve of Fig. 1 replotted in terms of some additional objective variables. A scale of nominal frequency band width required for transmission of the image signal over a video circuit has been worked out on the basis of our comparison of the out-of-focus motion picture with a television image of known characteristics, to be described in section 5. We see that in order to effect an increase in

sharpness which would be practically always discriminated under our experimental conditions, that is, a change of three or four liminal units, the frequency band would have to be increased from say 2.5 megacycles to about 4.5 megacycles. To effect an additional increase of the same

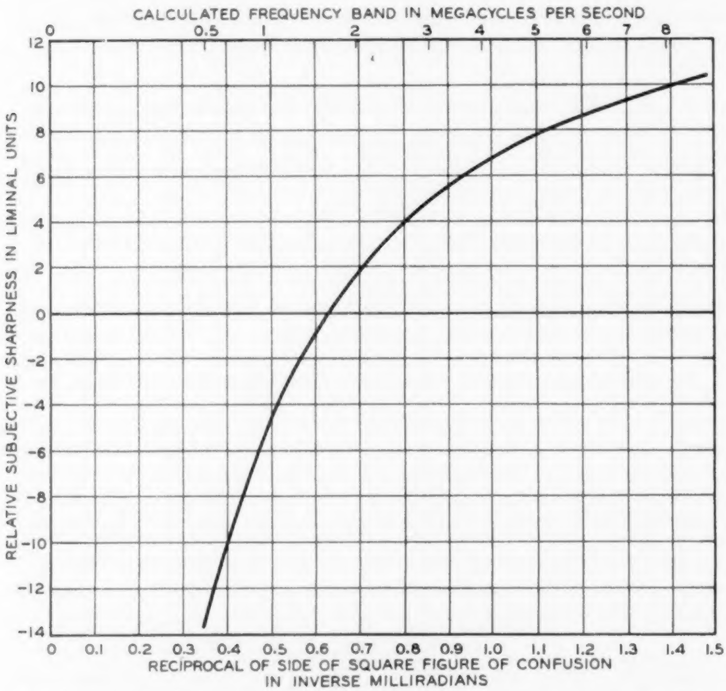


Fig. 4—Sharpness of small-sized motion pictures at a viewing distance of 30 inches. The frequency band is calculated on the same basis as in Fig. 3.

subjective amount would require that the frequency band be increased from 4.5 megacycles to about 10 megacycles. The diminishing return in sharpness is possibly better illustrated by the continually decreasing slope of the curve in Fig. 4, in which the abscissa is proportional to the square root of the frequency band, a factor which may perhaps be interpreted to represent roughly the cost of electrical transmission over a long system. We might infer from this curve that transmission costs are likely to increase faster than image sharpness, other things being equal.

4. Horizontal and Vertical Resolutions

The effect of unequal horizontal and vertical resolutions upon sharpness is shown in Fig. 5. The various rectangular figures of confusion, which were intercompared in a manner which will be described pres-

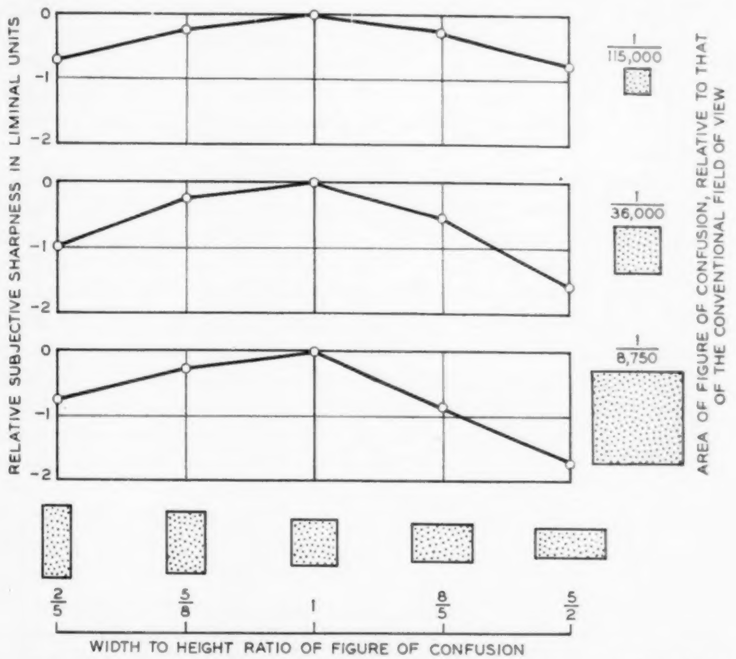


Fig. 5—Sharpness of small-sized motion pictures as a function of the relative values of horizontal and vertical resolutions. The conventional field of view is a rectangle whose height is 1/4 the viewing distance and whose width is 4/3 the height. Each point represents 150 observations at a viewing distance of 30 inches.

ently, are shown along the axis of abscissae, positioned according to the logarithm of the ratio of width to height, for the sake of symmetry. Three curves are shown, each for a different constant value of the area of the figure of confusion, which determines the sharpness for the central square shape (as in Fig. 1). At the right the relative areas are illustrated and specified in terms of the number of figures of confusion in the conventional field of view, whose width is 4/3 of its height and which is viewed at 4 times its height.

Sharpness, the subjective variable, is plotted along the axis of ordinates in liminal units. This unit denotes a difference in sharpness

which corresponds to a difference of one limen in the shape of the figure of confusion. When two images, characterized by different shapes of figure of confusion, are judged by a number of observers, the proportion of the observations in which one image is said to be sharper than the other affords a significant measure of the evaluation of the difference between them. When 25 per cent of the observations show that shape *A* yields a sharper image than shape *B*, we say that shapes *A* and *B* are different by one limen, and that the image *A* is less sharp than the image *B* by one liminal unit. All "no choice" or "equally sharp" judgments are distributed equally between the judgments for *A* and those for *B*.

In order to evaluate other than unit differences, we have assumed that a normal error curve describes accurately enough the distribution of sharpness differences in liminal units. Thus, image *A* is less sharp than image *B* by two liminal units when it is reported to be sharper than image *B* in 9 per cent of the observations. The difference is three liminal units when it is reported to be sharper in 2 per cent of the observations. Any difference larger than about three liminal units would indicate practically complete agreement that the one image is less sharp than the other, under our experimental conditions. A distribution of this nature was found to hold for sharpness differences resulting from changes in the area of the figure of confusion, as shown in Fig. 2.

Each shape of figure of confusion was compared with each of the four other selected shapes, and the sharpness differences were expressed in liminal units by the procedure just discussed. A fifth difference, corresponding to a null change, or a shape compared with itself, was presumed to be zero. The average value of these five sharpness differences, averaged in liminal units, measured the relative sharpness of that particular shape with respect to the average sharpness of all five shapes, an unvarying reference. In Fig. 5, the sharpness scales have been shifted so that zero denotes the most preferred one of the shapes, which happened in each case to be the square.

The sharpness curves are found to be slightly skewed with respect to the logarithm of the width : height ratio, there being a small preference for figures of confusion whose long dimension is vertical rather than horizontal. This is believed to be the first evidence of an asymmetric requirement on resolution. It suggests the possibility that the square figure might not have been the most preferred, had we tested other shapes nearer to the square than the ones we did use. With a more searching experiment we might have found that the eye prefers resolution in the horizontal direction to be just a little better than in

the vertical direction. Inasmuch as the effect is fairly small, and found only with the less sharp images, we shall leave it as another problem in physiological optics.

With an actual television image this small skewness would probably be reversed by the attendant coarsening of the scanning line structure. We do not know how much to allow for annoyance caused by visibility of the line structure. Taking our best estimate * of the height of the figure of confusion which would be equivalent in vertical resolution to a just noticeable pattern of scanning lines, we may say that for the uppermost curve in Fig. 5 the scanning line structure would not be noticeable except possibly for the shape marked 2/5. For the central curve the line structure would be noticeable for all shapes except possibly the one marked 5/2. It appears that the skewness and the line structure vanish together as the sharpness is increased.

Figure 5 demonstrates that equality of horizontal and vertical resolutions is a very uncritical requirement on the sharpness of an image, especially of a fairly sharp one. An image somewhat better than present television grade, exemplified by the uppermost curve in Fig. 5, shows a remarkably wide tolerance in this respect. Its figure of confusion could be three times as high as wide, or three times as wide as high, yet any intermediate shape between those two extremes would yield an equally sharp image to within one liminal unit. Under the ordinary conditions of television viewing the difference would be even less marked than that. This would imply that if the square figure of confusion simulates a television image of say 500 lines, then the number of lines could be changed to any value from about 300 to about 850 without altering the sharpness by as much as one liminal unit, under the condition, of course, that all the other pertinent factors, such as frequency band width and number of frames per second, remain unchanged.

The curves in Fig. 5 represent the averaged responses of fifteen observers each viewing five different motion picture scenes. Each one of the five selected shapes of figure of confusion was shown with each other one as a pair, a total of ten pairs. The observer was asked to identify which member of each pair he judged to yield the sharper image, or to report "no choice" if he judged them to be equally sharp. The pairs were scheduled in irregular order, and the observer could have the aperture shape shifted at will. The observers were instructed to consider the whole image area without undue regard for some features to the neglect of others.

* Engstrom [8] estimates that the scanning line structure becomes just noticeable when the spacing of the lines subtends an angle of 2 minutes at the observer's eye. In section 5 we show that the equivalent figure of confusion has a height 1.9 times as great as the spacing of the scanning lines.

5. Comparison of the Out-of-Focus Motion Pictures with a 240-Line Television Image of Known Characteristics

The motion picture machine was arranged to project out-of-focus pictures onto a screen set up beside the cathode ray receiving tube of a laboratory television apparatus [9] of excellent design. Duplicate films were run in the two machines, and the images were made equal in size and approximately equal in color and brightness. Special low-pass filters in the video circuit limited the frequency band without transient distortion, and permitted the trial of three different band widths. The conclusion was reached that the nominal band width of the video circuit, expressed in cycles per frame period, was equal to 1.3 times the number of figures of confusion in the frame area.

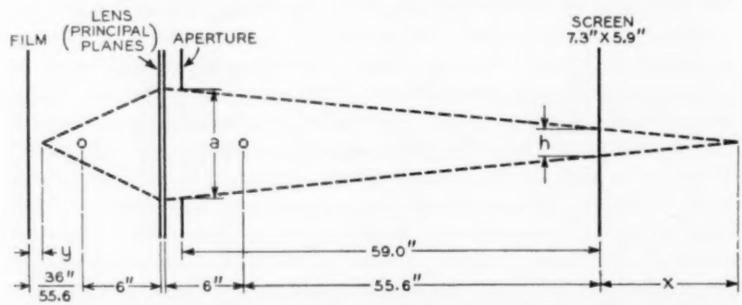


Fig. 6—Essential dimensions of the motion picture optical system as used for the correlation with a 240-line television image. For this case $a^2 = 1.00$ square inch.

$$y = \frac{36}{55.6} \cdot \frac{x}{55.6 + x}$$

$$\frac{h}{a} \doteq 1.45\, y.$$

A group of observers compared the two images, each observer being allowed to adjust the focus of the projection lens until he judged the images to be equal in sharpness. The distribution of lens positions, in terms of microscope scale divisions, was found to follow a normal error curve fairly well, and the median value for the group was used in computing the sizes of the figure of confusion. The external aperture shape was always square.

Since the television film scanner had been designed without regard for the unused space between frames on sound film, it became necessary to modify some of the dimensions of the out-of-focus projection system in order to make the two images equal in size. Figure 6 shows the modified dimensions. Comparison with Fig. 7, which gives the di-

mensions used in the main experiments, will show that the magnification was reduced from 12 to 9.3, the area of the external aperture was increased from 0.49 square inch to 1.00 square inch, and the aperture was mounted 2.6 inches instead of 1.3 inches from the principal planes of the projection lens.

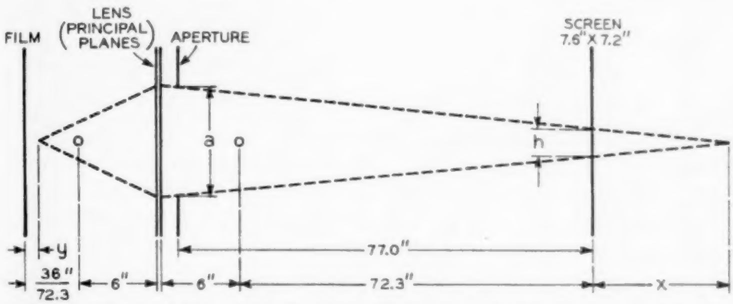


Fig. 7—Essential dimensions of the motion picture optical system as used for the subjective sharpness tests. For this case $a^2 = 0.49$ square inch.

$$y = \frac{36}{72.3} \cdot \frac{x}{72.3 + x}$$

$$\frac{h}{a} \doteq 1.88\, y.$$

The television apparatus was designed for 240 lines, 24 frames per second, and a width : height ratio of 7 : 6. Actually 20 per cent of the frame time was consumed in scanning the blank space between sound film frames, and 10 per cent of the line time was used up by the return sweep in the receiver. The television image, which was the same size as the projected motion picture, was 5.6 inches high. This dimension was 20 per cent less than the height of the entire 240-line field including the blank portion, which was, therefore, 6.9 inches. The width of the entire field including return trace was $6.9 \times 7/6$ or 8.1 inches, and the width of the television image was 10 per cent less than this, or 7.3 inches. Thus, the total area transmitted per frame period was 6.9×8.1 or 56 square inches; the useful image area was 5.6×7.3 or 41 square inches.

The three amplitude-frequency characteristics used in the video circuit are shown in Fig. 8. Curve A is for two square scanning apertures in tandem, one transmitting and one receiving, each having the height of one scanning line. No electrical band limitation was effective in this case. Curves B and C are for the addition of each of two special low-pass filters which were carefully phase-equalized and

designed for gradual cut-off. In each case the nominal band width was taken to be the same as that for the aperture effect alone, namely, the frequency at which the loss is 7.8 decibels greater than at low frequency. The addition of a low-pass filter could thus be considered

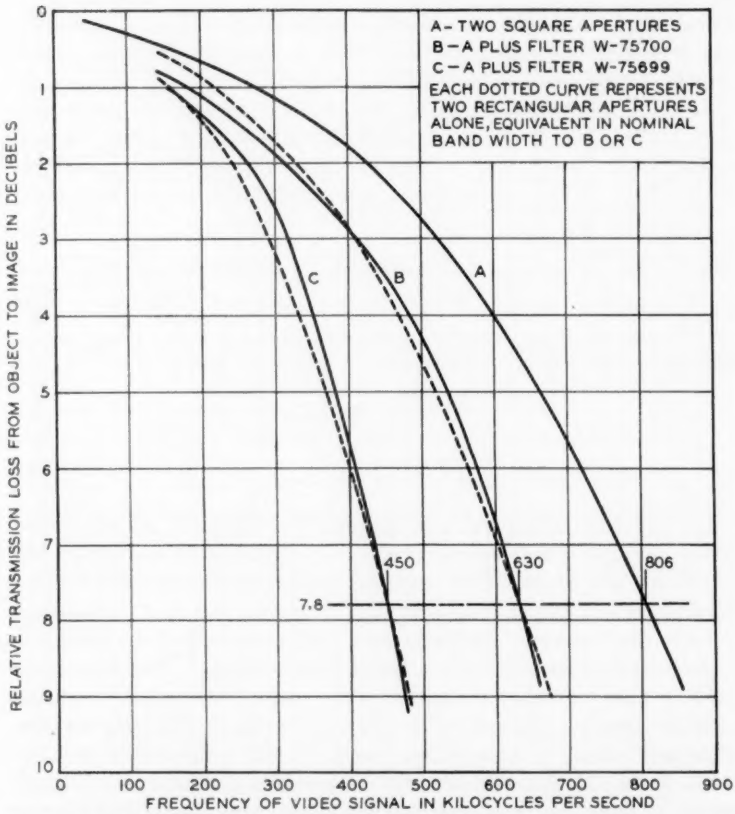


Fig. 8—Amplitude-frequency characteristics of the television system used for correlating the motion picture projection with a television image.

equivalent to an increase in the length of each of the scanning apertures in the ratio of the nominal band widths, as illustrated by the dotted curves.

The results of the comparison were as follows. The number of figures of confusion in the area of the frame was reckoned by dividing the entire area of the television frame, including the blank portion and

the return trace, by the observed area of one figure of confusion of the equally sharp motion picture. The number of cycles per frame period was the nominal band width of the video circuit, in cycles per second, divided by the number of frames per second, or 24.

	Case A	Case B	Case C
Figures of confusion per frame	22,400	18,900	14,800
Cycles per frame period	33,600	26,200	18,800
Ratio	1.50	1.38	1.27

The ratio in Case A was suspected to be too large because of unaccounted-for small defects in the film scanner which degraded the image sharpness more than was indicated by the aperture effect alone. The difference in ratio between Cases B and C was no larger than the measured probable error of each set of observations. Making allowance for these things, we concluded that the ratio between the number of cycles per frame period and the number of figures of confusion per frame area had been found to be 1.3.

This factor 1.3 gave us a basis for calculating the television aperture loss in the direction normal to the scanning lines, and enabled us to compute the nominal video frequency band required to yield an image having equal horizontal and vertical resolutions.

The stepped nature of the brightness variation across the scanning lines of a television image, in contrast to its continuous nature along the lines, gives rise to the requirement that for equal resolution in the two directions the scanning apertures must be longer in the scanning direction than they are across it. The extent of this departure from squareness has been estimated (see references 3 to 7) at from 1.2 to 1.9, mostly on theoretical grounds. Our comparison of a television image of known characteristics with a controlled out-of-focus motion picture furnished a subjective measurement of the effect which yielded the value 1.4 for the ratio of width to height of the scanning apertures for equal resolution. We take width to mean the dimension along the scanning lines, and height to mean the dimension normal to them.

We found that the nominal video band width of a television signal, in cycles per frame period, was 1.3 times the number of figures of confusion per frame area in the equally sharp motion picture. This meant that the area of each figure of confusion was 1.3 times as great as the area of one scanning line over a (scanned) length of one cycle. By the adopted definition of nominal video band width, the length of one cycle was just twice the length of each one of the pair of rectangular scanning apertures which were considered equivalent to the actual

square apertures plus the filter. According to the scanning theory of Mertz and Gray,[4] the pair of apertures in tandem was equivalent, in frequency limitation, to a single aperture 1.35 times as long as either one of the pair. Taking the width of this single aperture (1.35 times the length of one half cycle) equal to the width of the figure of confusion, the height of the figure of confusion was calculated from its area to be 1.9 times the height of one scanning line or one scanning pitch. This was the measure by subjective comparison of the resolution across the scanning lines.

Under the condition of equal resolution along and across the scanning lines, the figure of confusion would have to be square and its width would then also be 1.9 times the scanning pitch. The width of each one of the pair of equivalent tandem scanning apertures would be 1.9/1.35 or 1.4 times the scanning pitch. That is, two rectangular scanning apertures, each 1 line high and 1.4 lines wide, used in tandem without electrical band limitation, would yield an image having equal resolution along and across the scanning lines.

The nominal frequency band associated with such scanning apertures is 1/1.4 times that associated with square apertures. That is, the nominal video frequency band, in cycles per frame period, required for equal horizontal and vertical resolution is 0.70 times one half the number of square scanning elements per frame area, reckoning a square scanning element as an area of height and width equal to the scanning pitch, or spacing between scanning lines.

For comparison with the value 0.70 which we have just found, the following values of nominal band width coefficient have been lifted from their contexts in the references:

(*a*) Kell, Bedford and Trainer (1934)	0.64
(*b*) Mertz and Gray (1934)	0.53
(*c*) Wheeler and Loughren (1938)	0.71
(*d*) Wilson (1938)	0.82
(*d*) Kell, Bedford and Fredendall (1940)	0.85

Acknowledgment

This work has been done under the direction of Dr. P. Mertz and with the extensive assistance of Mr. T. R. D. Collins. To them, and to my other colleagues who have given of their time and counsel, I wish to extend my appreciation of their help.

APPENDIX

1. Determination of the Size of the Figures of Confusion

The image was put out of focus by moving the projection lens nearer to the film gate, throwing the plane of sharp focus beyond the viewing

screen. Assuming for the moment that the optical imagery was perfect, each point of the film gave rise to a pyramidal volume of light whose base was the opening of the external aperture and whose apex was the point's image in the new focal plane beyond the screen. The intersection of this pyramid with the viewing screen was the geometrical figure of confusion for that point. The shape of the figure was geometrically similar to that of the aperture, and the side of the figure was to the corresponding side of the aperture as the distance from focal plane to screen was to the distance from focal plane to aperture.

The distance of the focal plane beyond the screen was related to the displacement of the lens from the "in focus" position by means of the simple lens formula, and this relation was verified by actual measurement of the distances. The geometrical area of the figure of confusion was thus known in terms of the lens displacement, as shown in Fig. 9.

Efforts to check this relationship by direct measurement of the dimensions of the figure of confusion in the plane of the screen were nullified by the aberrations of the optical system, especially by the residual chromatic aberration. A comparison method was therefore devised in which the out-of-focus image of a very thin vertical slit was compared with an actual slit in the plane of the screen. In the film gate was placed a glass plate bearing a sputtered layer of gold with a razor-blade scratch not wider than 0.0001 inch in selected portions. In the plane of the screen was placed a back-lighted slit made by cementing the two halves of a cut piece of thin black paper onto a piece of translucent white paper. This slit had sharp, parallel edges and uniform brightness over its width, which was easily made as small as 0.005 inch. A set of these slits was prepared, ranging in width up to 0.100 inch, and each one was observed, without optical aid, close beside the projected out-of-focus image of the scratch in the gold film. The apparent brightnesses were equalized by means of neutral-tint filters behind the paper slit.

The ranges of values of lens displacement and of external aperture shape which were used in the experiments were tested in this way, by adjusting the out-of-focus images to subjective equality with the sharp-edged slits. In every case the measured width of the comparison slit turned out to be about 15 per cent less than the calculated geometrical width of the projected image. This seeming a not unreasonable measure of the effect of the aberrations, it was adopted as a factor for converting geometrical sizes into subjective sizes of the figures of confusion.

Figure 9 shows both the calculated geometrical area and the observed subjective area of the figure of confusion in terms of the displacement

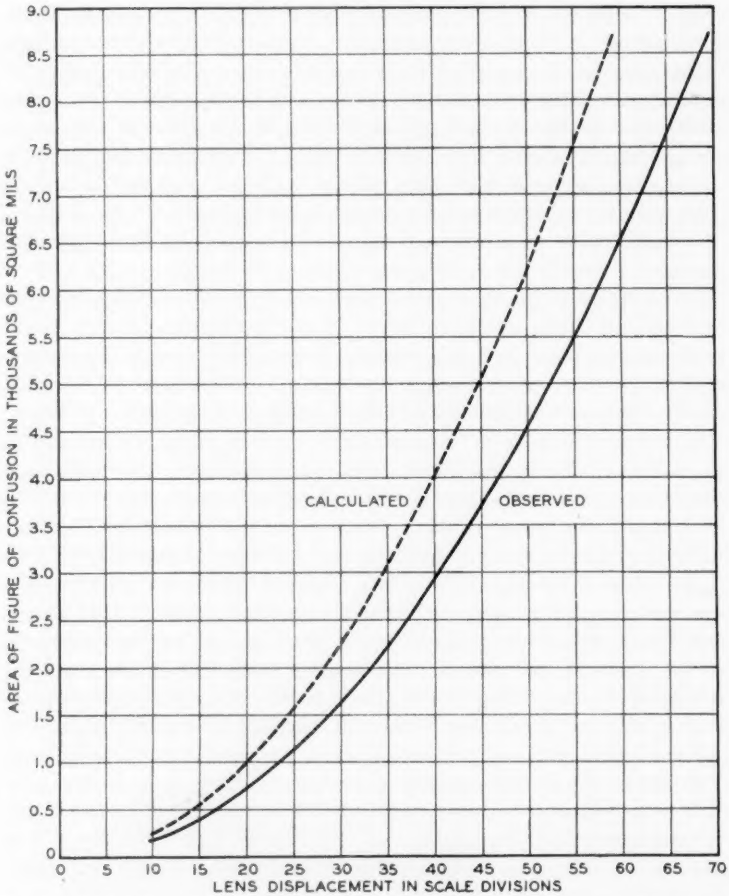


Fig. 9—Calibration curve for the motion picture optical system as used for the subjective sharpness tests.

of the lens from its sharp focus position. The lens displacement is expressed in microscope scale divisions, the working variable. Figure 7 shows the dimensions of the optical system.

2. Equipment and Conditions of the Experiment

Light Source

A ground glass screen ½ inch behind the film, illuminated by a 1,000-watt projection lamp and double condensing lens system. This served

to break up the image of the lamp filament which otherwise would have been formed near the principal planes of the projection lens and would have destroyed the uniformity of illumination within the figures of confusion of the out-of-focus image on the screen. The screen brightness was about 10 foot-lamberts with the projector running without film.

Projection Machine

Acme Portable, with two-bladed shutter. There was no provision for reproducing the sound track. The screen image, in sharp focus, was said by competent judges to represent very good motion picture projection practice.

Projection Lens

Bausch and Lomb Series "0", 6.00-inch focus. There was fitted over the lens barrel a brass ring with an extremely sharp turned edge to serve as an index for the measurement of lens displacement. The lens could be set to the nearest 0.0003 inch by means of the focusing mechanism. The image was put out of focus by moving the lens toward the film. At sharp focus the linear magnification was 12 times.

Measuring Microscope

Mounted rigidly on the frame of the projector, and fixed with respect to the film gate. The micrometer scale was focused on the index mark on the barrel of the projection lens. A lens displacement of 0.060 inch caused the index to traverse 50 divisions of the scale.

External Aperture

An adjustable black paper mask mounted 1⅓ inches from the principal planes of the lens, on the screen side. The opening was rectangular, with sides horizontal and vertical, of constant area 0.49 square inch. The ratio of height to width could be varied continuously from 2.5 to 0.40 without changing the area. The opening was uniformly filled with light under all conditions.

Viewing Screen

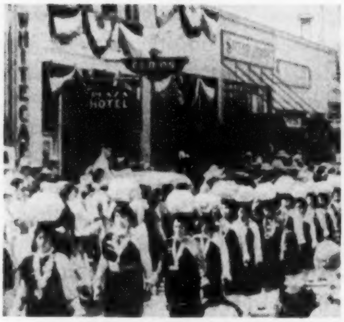
White Bristol board, 7.2 inches high by 7.6 inches wide (the image size of an available television receiver to be used for comparison). The screen was hung at the back of a black-velvet-lined box 18 inches high, 22 inches wide and 12 inches deep. The viewing distance was always 30 inches.

The viewing room was completely darkened except for a little stray light from the projection machine.

Scene 1

Scene 2

Scene 3



Scene 4

Scene 5

Fig. 10

Scene 1 reproduced by courtesy of Loucks & Norling.
Scenes 2 and 3 reproduced by courtesy of Fox Movietone News.
Scenes 4 and 5 reproduced by courtesy of Paramount News.

The Observers

The observers were almost all Laboratories engineers associated with television research and transmission problems. The average observer devoted about one hour to the experiment on unequal horizontal and vertical resolutions, and about three hours (in two sessions) to the experiment on small differences in resolution. Each observer was carefully instructed with regard to the purpose and the mechanism of the experiments, and was allowed to examine trial pictures to see clearly the effects of changing the shape and size of the figure of confusion.

Motion Picture Film

Standard 35 mm. black-and-white sound film on safety base. The area projected onto the screen was 0.600 inch high by 0.633 inch wide.

For the experiment on unequal horizontal and vertical resolutions, five different scenes were used. Sample frames from them are shown in Fig. 10. For the experiment on small resolution differences, Scene 3 was selected as the most suitable on the basis of photographic excellence and picture content, and this alone was used. Each of the scenes was about one quarter of a minute in length, and was shown repeatedly. Brief descriptions follow:

Scene 1: A country-side landscape, with trees and fields. A center of interest is the tall steeple of a white church on the distant hillside. A concrete highway flanked by a white fence carries cars into and out of the picture. There is no fast motion.

Scene 2: A full-length view of a girl modeling an evening dress moving slowly against a dark, fluted backdrop. A large vase of flowers is a secondary center of interest.

Scene 3: A close-up view of a girl modeling a hat, turning slowly against a plain, neutral background.

Scene 4: A street scene of an Indian parade, with a background of store windows and signs. The parade moves rather rapidly, and there is some motion among the by-standers.

Scene 5: A closer view of some of the Indians in the parade. There is much fine detail in the costumes, and the motion is rapid.

References

1. M. Luckiesh and Frank K. Moss, "The Dependency of Visual Acuity Upon Stimulus-Distance," *Jour. Opt. Soc. Amer.*, **23**, 25 (Jan. 1933).
2. J. P. Guilford, "Psychometric Methods," McGraw-Hill, New York, 1936.
3. R. D. Kell, A. V. Bedford and M. A. Trainer, "An Experimental Television System," *Proc. I.R.E.*, **22**, 1246 (Nov. 1934).

4. P. Mertz and F. Gray, "A Theory of Scanning and its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television," *Bell Sys. Tech. Jour.*, **13**, 464 (July 1934).
5. H. A. Wheeler and A. V. Loughren, "The Fine Structure of Television Images," *Proc. I.R.E.*, **26**, 540 (May 1938).
6. J. C. Wilson, "Channel Width and Resolving Power in Television Systems," *Jour. Telev. Soc.*, **2**, 397 (June 1938).
7. R. D. Kell, A. V. Bedford and G. L. Fredendall, "A Determination of Optimum Number of Lines in a Television System," *RCA Review*, **5**, 8 (July 1940).
8. E. W. Engstrom, "A Study of Television Image Characteristics," *Proc. I.R.E.*, **21**, 1631 (Dec. 1933).
9. H. E. Ives, "The Transmission of Motion Pictures over a Coaxial Cable," *Jour. S.M.P.E.*, **31**, 256 (Sept. 1938).

Cross-Modulation Requirements on Multichannel Amplifiers Below Overload

By W. R. BENNETT

Interchannel interference caused by non-linearity of multichannel amplifier characteristics is analyzed in terms of second and third order sum and difference products of the bands of energy comprising the various channels. Methods of relating the resulting disturbance to discrete frequency measurements are described and means for arriving at modulation requirements on individual amplifiers thus established.

1. Introduction

When a repeater is used to amplify a number of carrier channels simultaneously, departure from linearity in the response as a function of input amplitude tends in general to produce interference between the channels. The non-linear component of the amplifier characteristic in effect acts as a modulator, changing the frequencies in the input wave and producing components which fall in bands assigned to channels other than the original ones. This phenomenon has been called "interchannel modulation" or "non-linear crosstalk." In formulating the requirements which are imposed on a repeater to insure that the resulting interference between channels will not be excessive it is convenient to treat separately two aspects of the problem namely—the condition when the total load on the amplifier is within the range for which the amplifier is designed and the severely overloaded condition. Actually a transition region between these two cases must also exist but when a considerable amount of negative feedback is used the break in the curve of response vs. input is quite sharp so that for practical purposes the input may be said to be either below or above the overload value.

Below overload, the amplifier characteristic may in most cases be represented with sufficient accuracy by the first few terms of a power series and the interchannel modulation analyzed in terms of the resulting combination tones of the frequencies present in the different channels. The total interference resulting from the combination tones falling in individual channels must be kept below prescribed limits. Above overload on the other hand the resulting disturbance in all channels becomes quite large and requirements are based on making such occurrences sufficiently infrequent.

The load capacity requirement has been discussed in a paper by B. D. Holbrook and J. T. Dixon.[1] In the present paper we assume that the principles there developed have been used to fix the maximum load which the amplifier must deliver and proceed with the second phase of the problem—the determination of the modulation requirements which the amplifier should meet below overload. A secondary objective is to set up simple testing procedures by which the performance of the system may be assessed without incurring the complications attendant upon loading the system with talkers. One such procedure involves the measurement of distortion products by a current analyzer when sinusoidal waves are impressed upon the system. Another involves the measurement of noise in a narrow frequency band when a band of noise uniformly distributed over the transmission range is impressed upon the system.

The path followed in reaching these objectives starts in Section (2) with a demonstration of the way in which non-linearity leads to interference in specific cases. Section (3) then considers the magnitudes of typical modulation products and arrives at volume distributions for them, by which the fluctuating character of speech may be taken into account. The basis for treating interchannel modulation as noise is given in (4). Since one of the most convenient testing methods involves the use of sine waves the relationship of distortion measured with sine waves to the distortion observed with speech on the system must be set up as in (5). The number of products falling in any single channel is considered in (6) and the average noise in any channel may then be evaluated in (7) with the aid of the results of preceding sections. The effects introduced by multiple centers of distortion in the amplifiers of a transmission link are considered in (8). The paper concludes with a discussion of test methods presented in the ninth section.

2. Interchannel Modulation as a Source of Crosstalk

We shall consider specifically a single sideband suppressed-carrier multichannel speech transmission system of the four-wire type although much of the treatment is also applicable to other kinds of carrier systems. The carrier frequencies will be assumed to be adjacent harmonics of a common base frequency greater than the highest signal frequency. The amplifier characteristic will be assumed to be expressible with sufficient accuracy by means of linear square and cube law terms. That is if i_p represents the output current and e_g the input voltage we write

$$i_p = a_1e_g + a_2e_g^2 + a_3e_g^3. \tag{2.1}$$

[1] *Bell System Technical Journal*, Oct. 1939, Vol. 18, pp. 624–644.

Multivalued characteristics such as associated with ferromagnetic materials and reactive characteristics in which the coefficients vary with frequency are not included. The mechanism by means of which the characteristic (2.1) gives rise to interchannel interference may be illustrated by assuming that a sinusoidal signal of frequency q radians per second is impressed on the voice frequency channel associated with the carrier frequency mp, where p is the base frequency in radians per second and m is an integer. The resulting wave impressed on the amplifier is of the form

$$e_g = Q \cos (mp + q)t, \tag{2.2}$$

if upper sidebands are transmitted; the plus sign would be replaced by a minus sign in a system using lower sidebands. Substituting the value of e_g given by (2.2) in the characteristic (2.1), we find:

$$\begin{aligned} i_p = a_1 Q \cos (mp + q)t + \tfrac{1}{2}a_2 Q^2 + \tfrac{1}{2}a_2 Q^2 \cos (2mp + 2q)t \\ + \tfrac{3}{4}a_3 Q^3 \cos (mp + q)t + \tfrac{1}{4}a_3 Q^3 \cos (3mp + 3q)t. \end{aligned} \tag{2.3}$$

Considering the terms which appear in the response (2.3), we note that the first term is the desired amplified signal. The second term is a direct current of trivial importance; if the system contains a transformer, this component is not transmitted. If q does not exceed $p/2$, the third term will be received in the channel associated with the carrier frequency $2mp$ and will there produce a detected frequency twice as great as the original signal frequency. In such a case, it represents interference produced in the $2mp$-channel when the mp-channel is actuated. If q exceeds $p/2$, the interference falls in the $(2m + 1)p$-channel. The fourth term of (2.3) is received in the mp-channel and represents a non-linearity in intrachannel transmission since its frequency is the same as that of the applied signal but its amplitude is proportional to the cube of the impressed signal amplitude. This term is of trivial interest in the study of transmission quality of individual channels of a well-designed multichannel system and is of no interest in the interference problem with which we are here concerned because it is received only in the originating channel. Finally, if q is less than $p/3$, the fifth term represents interference of frequency $3q$ received in the channel associated with the carrier frequency three times that of the originating channel; if q is greater than $p/3$, the interference falls in a higher channel.

Next suppose that a number of carrier channels are simultaneously transmitting signals. By substituting an expression representing the resulting carrier wave, which is a sum of several terms such as (2.2) with different values of Q, m and q, in the amplifier characteristic (2.1),

we find that in addition to interference in channels having twice and three times the fundamental carrier frequencies, there are modulation terms appearing in channels having carriers which are various combinations of sums and differences of the original carrier frequencies. The second order term gives rise to crosstalk products with carriers $(m + n)p$ and $(m - n)p$ as well as $2mp$. The third order term causes products with carriers $(2m + n)p$, $(2m - n)p$, $(l + m + n)p$, $(l + m - n)p$, and $(l - m - n)p$ as well as $3mp$. In the above l, m and n are integers. For convenience we represent the tones associated with the carriers lp, mp, np by A, B and C and designate the various types of products as $A + B$, $A - B$, $2A + B$, $2A - B$, $A + B + C$, etc. It will be noted that the resultant modulation falling in a particular channel at any instant depends on the particular loading conditions prevailing on other channels at the same instant, and that a wide variety of amplitudes, numbers, and types of products are possible. Detailed study of these possibilities is necessary for the solution of our problem.

3. Nature of Multichannel Speech Load and Resulting Modulation Products

Considering any individual channel of the system, we note that (1) it may be active or inactive and (2) if active, the signal power being transmitted may vary throughout a considerable range of values. With regard to (1), we may estimate from traffic data a probability τ that a channel is active.[2] With regard to (2), data are available on the distribution of volumes corresponding to different calls at the toll switchboard. By "volume" is meant the reading of a volume indicator of a standard type. The distribution is approximately normal and hence may be expressed in terms of the average value V_0 and standard deviation σ. In mathematical language, the probability that the volume from any subscriber is in the interval dV at V is given by

$$p(V)dV = \frac{1}{\sigma\sqrt{2\pi}} e^{-(V-V_0)^2/2\sigma^2} dV. \tag{3.1}$$

The value of V_0 is about 16 db below reference volume, or about -8 vu when measured on the new volume indicator recently standardized in the Bell System. The value of σ is about 6 db. It is to be noted that volume is proportional to the logarithm of average speech power and hence V_0 is not the volume corresponding to the mean of the average speech powers of different talkers. The latter quantity, which will

[2] A channel is said to be active whenever continuous speech is being introduced into it. See Reference (1).

be designated by V_{0p}, may be calculated by averaging the distribution according to power, thus

$$\begin{aligned} V_{0p} &= 10 \log_{10} \overline{\text{antilog}_{10} V/10} \\ &= V_0 + \frac{\sigma^2}{20} \log_e 10 \qquad (3.2) \\ &= V_0 + .115\,\sigma^2, \end{aligned}$$

when V_0 and σ are expressed in db. The method of obtaining this result is indicated in Appendix A.

It will be convenient to extend the term "volume" to apply to modulation products by designating the modulation product produced by 0-vu talkers as a "zero volume modulation product" of its particular type. This is not its absolute volume as read by a volume indicator, but a reference value to which modulation products of the same type produced by talkers of other volumes may be referred. We assume on the basis of a power law of modulation that the volume of a product will increase one db for each one db increase in volume of a fundamental appearing once in the product, two db for each db increase in volume of a fundamental appearing twice, etc. Thus a $(2A - B)$-product should increase two db for one db increase in the volume of the A-component, and one db for one db increase in the volume of the B-component. If the fundamental talker volumes producing a particular product are normally distributed on a db scale, it follows from established relations concerning the distributions of sums [3] of normally distributed quantities that the volume of the product is also normally distributed. The relations between average and standard deviation for the modulation product and the corresponding quantities V_0 and σ for the fundamental are shown in Table I.

TABLE I

Modulation Product	Average in db Referred to Product from 0-vu Talkers	Standard Deviation in db
$2A$	$2V_0$	2σ
$A \pm B$	$2V_0$	$\sqrt{2}\sigma$
$3A$	$3V_0$	3σ
$2A \pm B,\ B - 2A$	$3V_0$	$\sqrt{5}\sigma$
$A + B \pm C,\ A - B - C$	$3V_0$	$\sqrt{3}\sigma$

That is, if the fundamental talker volumes are normally distributed with average value -8 vu, and standard deviation 6 db, the

[3] Multiplying amplitudes of fundamental components is equivalent to adding logarithms of amplitudes; hence the volumes of the fundamental components add to determine product volumes. For a derivation of the distribution function of the sum of two independent normally distributed quantities, see Cramer, Random Variables and Probability Distribution, Cambridge Tract No. 36, 1937, p. 50.

$(A + B - C)$-type products, for example, are also normally distributed in product volume with average value 24 db less than the product produced by three 0-vu talkers and standard deviation $6\sqrt{3}$ or 10.4 db. To obtain the product volume corresponding to the average power of the product distribution, $.115(10.4)^2$ or 12.4 db must be added. In general if V_{0p} of an x-type product is desired, it may be expressed as $\eta_x V_0 + .115\lambda_x\sigma^2$ where η_x is the order of the x-type product, and the value of λ_x is given by the square of the coefficient of σ in the third column of Table I.[4] We observe that $\eta_x V_0 + .115\lambda_x\sigma^2 = \eta_x V_{0p} + .115(\lambda_x - \eta_x)\sigma^2$, and that $\lambda_x = \eta_x$ for $x = A \pm B$ and $A \pm B \pm C$.

The frequencies present in a typical commercial speech channel extend over a range of approximately 3000 cycles. The spacing of carrier frequencies must be made somewhat greater than this to allow for filter cut-offs. Figures 1 and 2 illustrate the spectra of the various second and third order modulation products resulting from two and three fundamental channel spectra respectively which are flat from 10 per cent to 80 per cent of the carrier spacing. Actual speech channels would have peaked spectra but the results would be roughly similar. Each second order band of products occupies twice the frequency range of one original speech band, and a third order band of products spreads over three times the fundamental range. Portions of one product band may thus be received in different channels, but with one part usually much larger than the others. It is to be noted that a $2A$-type product band does not consist merely of the second harmonics of all tones in the band A, but includes all possible sums of the tones in the fundamental band. The spectrum of the $2A$-type product is similar in shape to that of an $(A + B)$-type product but has half as much total power because only half as many sum products can be formed from a single band as from two equal bands. The interfering effect of a $2A$-type product from a speech channel may of course be quite different in character from that of an $(A + B)$-type product since in the latter case the result depends on two independent talkers.

4. The Noise Resulting from Modulation Products

It will be noted that the interference produced as described above may be classified as unintelligible, since in products involving one channel, the wave form is distorted, and in products involving more than one channel, sums and differences of independent signal frequencies are heard. It may be said therefore that interchannel modulation

[4] In general for a $(m_1A \pm m_2B \pm m_3C \pm \cdots)$-product, $\lambda_x^2 = m_1^2 + m_2^2 + m_3^2 + \cdots$.

may be treated as noise, and the usual noise requirements apply. If a great many products are superimposed, the noise heard will be fairly steady, and the average weighted noise power is a sufficient indication of the interfering effect. In systems with a small number of channels, large variations in the noise may occur, and it may be necessary to consider the infrequent large bursts of modulation from exceptionally loud talkers as a limiting factor. The allowance to be made can be estimated by determining the complete distribution curve of modulation noise. Computation of the required distribution function may be carried out by methods similar to those described in the paper by Holbrook and Dixon.[1] The fact that the products are not independent introduces a difficulty which complicates the calculation. For systems with a large number of channels, the requirements may be based on average values with a considerable resulting simplification.

If in addition to the sidebands due to speech, "carrier leaks" (partially suppressed carrier waves) are present, modulation products are produced which are sums and differences of carrier frequencies and speech sidebands. Products of this sort may cause intelligible crosstalk. For example the carrier frequency mp modulating with the channel frequency $np + q$ causes an $(A + B)$-product of frequency $(m + n)p + q$, which is received in the channel with carrier frequency $(m + n)p$ as the original signal frequency q. Requirements on intelligible crosstalk are in general more severe than on unintelligible; hence it is important that the carrier leaks be suppressed well below the level of the speech channels. The intelligibility tends to disappear as the number of channels is increased, since the number of superimposed products becomes larger thereby producing masking effects. Carrier leak modulation is however more serious than modulation from speech channels having the same power since carrier leaks are present all the time, while speech sidebands occur only in active channels. Similar considerations apply to pilot and control tones.

Quantitatively, the various frequency components in modulation noise must be weighted in terms of their interfering effect on reception of speech. In practice the weighting is done by a noise meter designed for that purpose. The noise meter readings are expressed in terms of db above reference noise. A reading of zero, or reference noise, is produced by a 1000-cycle sinusoidal wave with mean power equal to one micromicrowatt. The weighting incorporated in the noise meter is determined by judgment tests of the relative interfering effects of single frequencies and other reproducible noises.

5. Relation Between Speech and Sine Wave Modulation

A goal of our investigations is to express the requirements finally in terms of measurements which can be made on amplifiers with sinusoidal testing waves. A means of relating modulation products produced by speech channels to those occurring when discrete frequencies are applied is therefore needed. For our purposes here we shall express the needed relation [5] in terms of a "Speech-Tone Modulation Factor," which we shall abbreviate as S.T.M.F. and define in terms of the following procedure: Apply the fundamental single-frequency test currents necessary to produce the product in question, which we shall designate as an x-type product. Adjust each fundamental to give mean power of one mw. at the zero level point of the system. Measure the resulting x-type product at the point of zero transmission level of the system. Suppose the product is H_x db below one fundamental. Next load the system with speech from the combination of fundamental talkers required to form the talker product of x-type. Each talker must produce speech volume of 0 vu at the transmitting toll switchboard or point of zero transmission level. The product is then received from the appropriate channel and a comparison is made between it and the speech from one talker with both talker and product at the same transmission level point in the system. The comparison should be made on the basis of relative interfering effect. Suppose it is determined that an x-type product is L_x db below one 0-vu talker. Then s_x, the S.T.M.F. for an x-type product, is defined by

$$s_x = L_x - H_x. \tag{5.1}$$

The sign of the S.T.M.F. has been assigned here to be positive when the difference in db between effect of talker and talker product is greater than the difference between sine wave fundamental power and sine wave product power.

It is to be noted that not only does each type or product possess its own S.T.M.F., but also that the several portions of a product appearing in different channels have different S.T.M.F.'s. This may be clearly seen from Figs. 1 and 2. We note also that the S.T.M.F. is a property of the system on which the measurements are made, since it varies with the band width of the channels, the spacing of carrier frequencies, and the extent of departures from the simple square and cube law representation of the amplifier characteristic. It also varies with the type of transmitting and receiving instruments used. Theo-

[5] The quantity here defined is related to what has been called "staggering advantage" of modulation products. Since the term "staggering advantage" has been applied to various kinds of interference including linear crosstalk, its use here might lead to confusion and is avoided.

retically it should be possible to compute the S.T.M.F. for any particular product if sufficient information concerning the properties of speech, the transmitting and receiving instruments, the carrier system

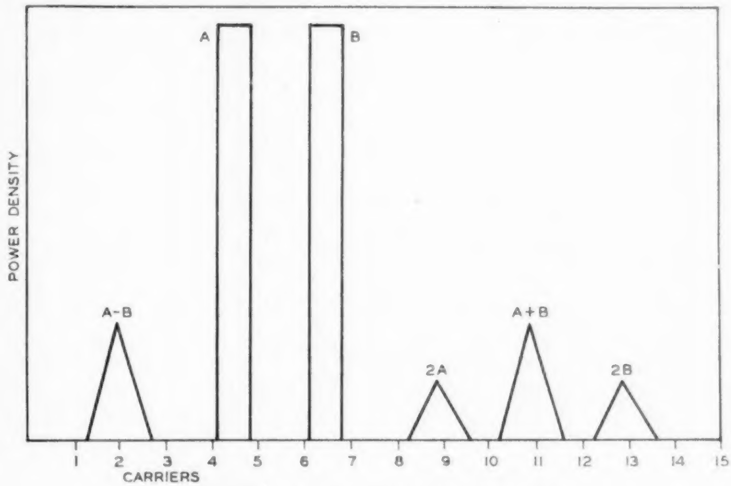


Fig. 1—Spectra of second order modulation products from two fundamental channels.

itself, and the ear were known, but in practice it is found best to use experimental determinations. In the case of new systems for which experimental data are not available, estimates based on known systems of similar type would be used.

6. Number of Products Falling in Individual Channels

In the appendix it is explained how the total number of possible products of each type falling in individual channels may be counted. Table II shows the result of counting all second and third order type products.[6] Results for products falling both within and without the fundamental band are given. In certain of the $(A + B - C)$- and $(A - B - C)$-type products, the channel in which the product occurs also is the source of one of the fundamentals. Since this would give a type of interference heard only when the disturbed channel is also carrying signal, it is not in general as serious a form of crosstalk as the cases of independent fundamental and product frequencies. Therefore the number of these special kinds of products has also been evaluated

[6] Mr. J. G. Kreer collaborated in the derivation of these formulæ.

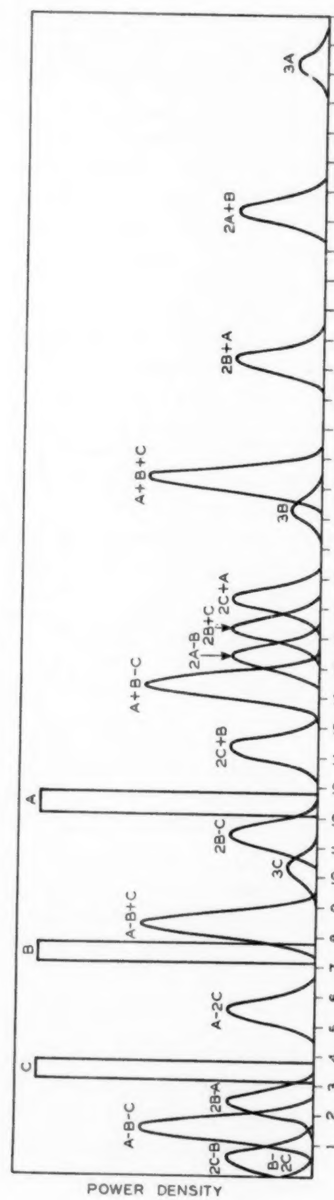


Fig. 2—Spectra of third order modulation products from three fundamental channels.

TABLE II

NUMBER OF PRODUCTS FALLING IN kTH CHANNEL OF MULTICHANNEL CARRIER SYSTEM WITH HARMONIC CARRIER FREQUENCIES n_1p, $(n_1+1)p$, $\ldots$ $(n_1+N-1)p$

n_1p = Lowest Carrier Frequency.
$n_2p=(n_1+N-1)p$ = Highest Carrier Frequency.
kp = Carr. Freq. Associated with Mod. Product.
N = Number of Channels.
$I(x)$ = "Largest Integer $\leq x$."
No. of Products is 0 Outside Ranges Indicated.

NUMBER OF PRODUCTS

Type

Second Order

$2A$........ $1,\ n_1\leq\frac{k}{2}\leq n_2$ if $\frac{k}{2}$ is an integer

$$A+B\quad\begin{cases} I\left(\frac{k+1}{2}\right)-n_1,\ 2n_1-1\leq k\leq n_1+n_2\\ n_2-I\left(\frac{k}{2}\right),\ n_1+n_2\leq k\leq 2n_2\end{cases}$$

$A-B$..... $N-k,\ 0<k<N$

Third Order

$3A$....... $1,\ n_1\leq\frac{k}{3}\leq n_2$ if $\frac{k}{3}$ is an integer

$$2A+B\ldots\begin{cases} I\left(\frac{k-n_1}{2}\right)-I\left(\frac{k}{3}\right)+I\left(\frac{k-1}{3}\right)-n_1+1,\ 3n_1\leq k\leq 2n_1+2n_2\\ I\left(\frac{k-n_1}{2}\right)-I\left(\frac{k}{3}\right)+I\left(\frac{k-1}{3}\right)-I\left(\frac{k-n_2+1}{2}\right)+1,\ 2n_1+n_2\leq k\leq n_1+2n_2\\ n_2-I\left(\frac{k}{3}\right)+I\left(\frac{k-1}{3}\right)-I\left(\frac{k-n_2+1}{2}\right)+1,\ n_1+2n_2\leq k\leq 3n_2\end{cases}$$

$$2A-B\ldots\begin{cases} I\left(\frac{n_2+k}{2}\right)-n_1+1,\ 2n_1-n_2\leq k\leq n_1-1,\ k\geq 0\\ I\left(\frac{n_2-k}{2}\right)-I\left(\frac{n_1+k+1}{2}\right),\ n_1\leq k\leq n_2\\ n_2+1-I\left(\frac{n_1+k+1}{2}\right),\ n_2+1\leq k\leq 2n_2-n_1\end{cases}$$

$A-2B$.... $I\left(\frac{n_2-k}{2}\right)-n_1+1,\ 0<k\leq n_2-2n_1,\ n_2\geq 2n_1$

$$A+B+C\ldots\begin{cases} I\left(\frac{k-3n_1+3}{6}\right)+I\left[\frac{(k-3n_1-1)^2}{12}\right],\ 3n_1+3\leq k\leq 2n_1+n_2+1\\ (N-1)\left[k-2n_1+1-\frac{N}{2}\right]+I\left(\frac{3n_2-k+3}{6}\right)+I\left[\frac{(3n_2-k-1)^2}{12}\right]-I\left[\frac{(k-3n_1)^2}{4}\right]\\ \qquad -\frac{1}{2}I\left(\frac{k-3n_1+2}{3}\right)I\left(\frac{k+3n_1+5}{3}\right)-\frac{1}{2}I\left(\frac{3n_2-k}{3}\right)I\left(\frac{3n_2+k+5}{3}\right),\\ \qquad\qquad 2n_1+n_2+2\leq k\leq n_1+2n_2-2\\ I\left(\frac{3n_2-k+3}{6}\right)+I\left[\frac{(3n_2-k-1)^2}{12}\right],\ n_1+2n_2-1\leq k\leq 3n_2-3\end{cases}$$

$$A+B-C\ldots\begin{cases} I\left(\frac{N-n_1+k}{2}\right)I\left(\frac{N-n_1+k+1}{2}\right),\ 2n_1-n_2+1\leq k\leq n_1-1\\ I\left(\frac{k-n_1}{2}\right)I\left(\frac{k-n_1-1}{2}\right)+(k-n_1)(n_2-k)+I\left(\frac{n_2-k}{2}\right)I\left(\frac{n_2-k-1}{2}\right),\ n_1\leq k\leq n_2\\ \qquad \text{Note: Number of products included in which } k\text{-channel signal is one fund. component}=\begin{cases} k-n_1,\ n_1\leq k\leq n_1+\frac{N-1}{2}\\ n_2-k,\ n_1+\frac{N-1}{2}<k\leq n_2\end{cases}\\ \left[n_2-I\left(\frac{k+n_1}{2}\right)\right]\left[I\left(\frac{k+n_1}{2}\right)-k+N\right],\ n_2<k\leq 2n_2-n_1-1\end{cases}$$

$A-B-C$.. $I\left(\frac{N-k-n_1}{2}\right)I\left(\frac{N-k-n_1-1}{2}\right),\ 0<k\leq n_2-2n_1-1,\ n_2\geq 2n_1+1$

$$\text{Note: Number of products included in which } k\text{-channel signal is one fund. component}=\begin{cases} N-2k-1,\ n_1\leq k\leq\frac{n_2-1}{3}\\ k-n_1,\ \frac{n_2-1}{3}\leq k\leq\frac{n_2+1}{3}\\ N-2k,\ \frac{n_2+1}{3}\leq k\leq\frac{N}{2}.\end{cases}$$

and shown in the table. The number of these is to be subtracted from the total of the $(A + B - C)$- or $(A - B - C)$-types to obtain the number of products not involving the listening channel. It should be pointed out also that kp is the derived carrier frequency throughout and that the products may extend over into adjacent channels. The principal component of the product usually falls in the channel with carrier frequency kp, but in some cases the amount of energy falling in adjacent channels may be quite considerable, as may be seen from Figs. 1 and 2.

The average number of products falling simultaneously in one channel is found by multiplying the total possible number by τ^2 for two-frequency products and τ^3 for three-frequency products, these factors being the probability that any particular product is present. The average number present is not affected by the dependence of the products arising from the fact that one talker may participate in the formation of more than one of the products falling in a channel. For convenience in making use of the results of Table II in evaluating the amplifier requirements, we shall represent the number of x-type products falling in channel number k when it is idle and all other channels are active by the symbol ν_{xk}. We shall also let $\mu(x)$ represent the number of distinct fundamental components required to produce an x-type product, e.g., $\mu(A + B) = 2$, $\mu(2A + B) = 2$, $\mu(A + B - C) = 3$, etc. It follows that the probability that any particular product is present is $\tau^{\mu(x)}$, since τ is the probability that any one required component is present. The average number of x-type products present in the k-channel is therefore $\nu_{xk}\tau^{\mu(x)}$, and may be considered as determined since ν_{xk} is the quantity tabulated in Table II.

7. Modulation Requirement in Terms of Average Total Noise Permissible in a Channel

From Section 3 we have a result for the volume of one product of arbitrary type, averaged on a power basis for a distribution of fundamental talker volumes, referred to the product of the same type produced by zero volume talkers. From Section 6 we have the average number of products of each type appearing in a channel. Combining these two results should give the average total modulation of each type present in a channel. A difficulty occurs however inasmuch as it is not certain how the interfering effect of superimposed modulation adds. The noise caused by one modulation product is of an irregular nature and it is probable that its most disturbing effect is associated with infrequent peak values. When two products are superimposed their individual peaks are not apt to coincide and hence the resultant dis-

turbance may not be much greater than that of one alone. We shall introduce here the concept of "plural S.T.M.F." Suppose ν products of x-type are superimposed and comparison of the resulting noise with one fundamental talker shows that the difference is $L_{x\nu}$ db. If interfering effects add as mean power we should expect $L_{x\nu}$ to be equal to $L_x - 10 \log_{10} \nu$. Hence it seems logical to write

$$s_{x\nu} = L_{x\nu} + 10 \log_{10} \nu - H_x, \tag{7.1}$$

where $s_{x\nu}$ is the "plural S.T.M.F." to be used when ν products are superimposed in order that power addition of products may be valid. Combining (5.1) and (7.1),

$$L_{x\nu} - L_x = s_{x\nu} - s_x - 10 \log_{10} \nu \tag{7.2}$$

which shows that the correction to be subtracted from power addition is

$$\rho_{x\nu} = s_{x\nu} - s_x. \tag{7.3}$$

The value of $\rho_{x\nu}$ is best determined by experiment. Superposition of a large number of products without using an excessive number of talkers can be accomplished by making phonograph records of individual products and combining their outputs in subsequent recordings.

The average total modulation of x-type in a channel is found by multiplying the average value for one product by the average number of products, and subtracting the quantity $\rho_{x\nu}$, which may be called the "plural S.T.M.F. correction," thus

$$V_x = \eta_x V_{0p} + .115(\lambda_x - \eta_x)\sigma^2 + 10 \log_{10} \nu_{xk}\tau^{\mu(x)} - \rho_{x\nu}. \tag{7.4}$$

where V_x is the volume averaged on a power basis of the x-type modulation in the k-channel referred to the volume of one x-type product from 0-vu talkers. We next wish to express V_x in terms of db above reference noise.

Let T_a represent the "noise" produced by a 0-vu talker in db above reference noise. This is an experimentally determinable quantity and is about 82 db. Let T_x represent the noise from an x-type product from 0-vu talkers. Then L_x, the quantity appearing in (5.1), is given by

$$L_x = T_a - T_x. \tag{7.5}$$

The average total noise produced by all x-type products in db above reference noise is given by

$$W_x = V_x + T_x = V_x + T_a - s_x - H_x. \tag{7.6}$$

If we assume that the total modulation noise allowable for an x-type product is X db above reference noise at zero level, we may equate X to W_x in (7.6) and solve for H_x, giving

$$H_x = V_x + T_a - s_x - X. \tag{7.7}$$

Substituting the value of V_x from (7.4) in (7.7), we get for the system requirement in terms of allowable ratio of fundamental to product when there is one mw. of each fundamental at the point of zero transmission level:

$$H_x = T_a + \eta_x V_{0p} - s_x + .115(\lambda_x - \eta_x)\sigma^2 + 10 \log_{10} \nu_{xk} + 10\mu(x) \log_{10} \tau - \rho_{x\nu} - X. \tag{7.8}$$

For the convenience of the reader, the following recapitulation of significance of the symbols used in (7.8) is given:

H_x = Ratio in db of power of each single frequency fundamental to power of resulting x-type product when each fundamental has power of one mw. at point of zero transmission level.

T_a = Reading of 0-vu talker on noise meter in db above reference noise.

η_x = Order of x-type product.

V_{0p} = Volume in vu corresponding to the average power of the talker volume distribution at the point of zero transmission level $= V_0 + .115\sigma^2$ where V_0 is average talker volume in vu.

σ = standard deviation in db of talker volume distribution.

s_x = Speech-Tone Modulation Factor of x-type product as defined in Section 5.

λ_x = $\sqrt{m_1^2 + m_2^2 + \cdots}$ for $(m_1A \pm m_2B \pm \cdots)$-type product.

ν_{xk} = total number of x-type products which can fall in channel with carrier frequency kp. See Table II.

$\mu(x)$ = number of distinct fundamentals required to produce x-type product.

τ = fraction of busiest hour that a channel is active.

$\rho_{x\nu}$ = correction to be applied to S.T.M.F. when ν products are superimposed. Defined in (7.1)–(7.3).

X = Allowable modulation noise in channel in db above reference noise at zero transmission level point.

The allowable noise may be divided equally between second and third order purely on a power basis by setting the requirement for each 3 db more severe than the total value allowed, or it may turn out that the noise from one order is much more difficult to reduce than that of the other in which the full allowance may be given to the more

difficult one, and the other made to contribute a negligible amount. Usually one type of modulation product will predominate for a given order and the allowable noise for the order may be assigned to this type. If such is not the case division of the noise between the various types may be estimated.

8. Addition of Products in a Multirepeater Line

When a number of amplifiers are used in a carrier system, contributions to interchannel interference occur at each repeater. The considerations previously discussed have a bearing on the modulation requirements on the system as a whole, but it is evident that the individual amplifiers may have to meet much more severe requirements. The relation between total system modulation and single amplifier modulation depends to a considerable extent on the phase angles between products originating in the various repeater sections.

A discussion of the general problem of addition of modulation products from multiple sources is to be given in a forthcoming paper by J. G. Kreer. A point of particular interest in connection with broad band systems is the effect of a linear phase characteristic on the phase shift between modulation products originating in different repeater sections. The curve of phase shift vs. frequency throughout the frequency range occupied by a considerable number of adjacent channels will in general depart but little from a straight line, but the intercept of this straight line if produced to zero frequency is in general not zero or a multiple of 2π. The intercept of such a linear phase curve is effective in producing phase difference between contributions to modulation from successive repeater sections of all the second order products and of some of the third order products, namely the types $3A$, $2A + B$, $B - 2A$, $A + B + C$, and $A - B - C$. The phases of third order products of types $2A - B$ and $A + B - C$ however are unaffected by the value of the intercept and the contributions from the different repeaters of these types of modulation will add in phase to give the maximum possible sum whenever the phase curve is linear throughout the channels involved. Third order modulation requirements on individual repeaters of a system may therefore have to be based on the very severe condition of in-phase or voltage addition of separate contributions.

Experimental verification of in-phase addition of third order modulation products from the repeaters of a 12-channel cable carrier system are included in the paper by Kreer previously mentioned. Corroborating data obtained on an experimental system capable of handling 480 channels are shown in Fig. 3. The measurements there shown were

made [7] on a loop approximately 50 miles in length with repeaters spaced 5 miles apart. The band transmitted extended from 60–2060 kc. Fundamental frequencies of 920 and 840 kc. were supplied from

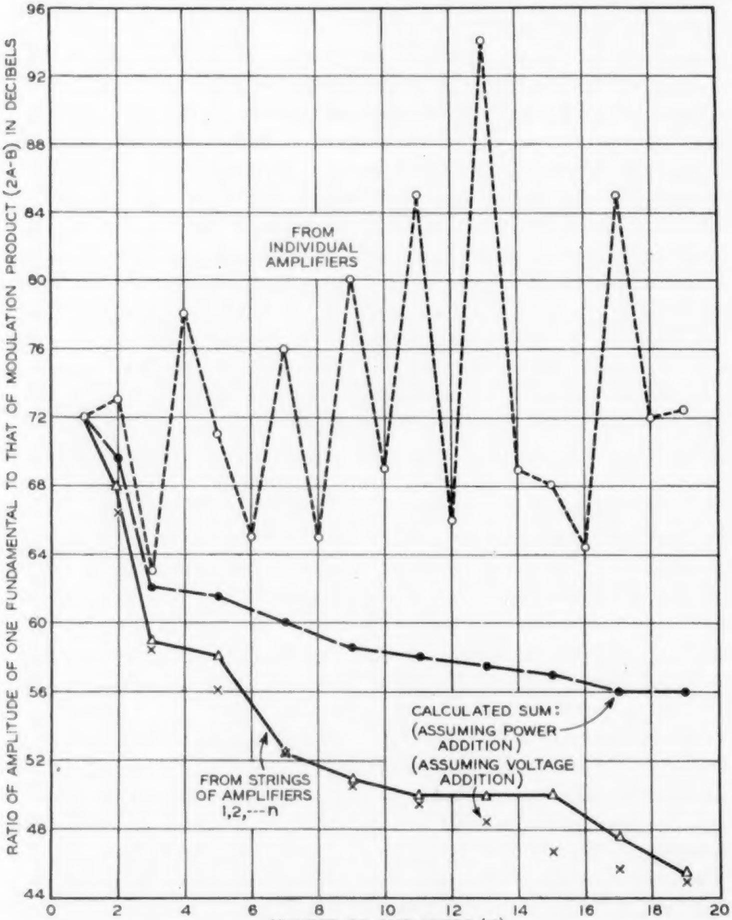


Fig. 3—Experimental data on addition of third order modulation from a multi-repeater line. Fundamental test tones $A = 920$ kc, $B = 840$ kc. Modulation product $2A-B = 1000$ kc.

two oscillators at the sending end, and measurements with a portable current analyzer were made at each repeater to determine the ratio of

[7] Messrs. M. E. Campbell and W. H. Tidd collaborated in these measurements.

amplitude of fundamental to that of the $(2A - B)$-product falling at 1000 kc. A band elimination filter having more than 100 db loss at 1000 kc. and suppressing a band approximately from 940 to 1070 kc. was inserted in the line at the repeater station to remove all contributions to the modulation product originating ahead of the station at which measurements were made. In this way, the modulation contributed by each amplifier and by various combinations of amplifiers could be measured without disturbing the operating levels throughout the system.

The data shown on Fig. 3 include measured modulation from individual amplifiers, and from tandem amplifiers with intervening cable sections. The summation of amplifiers proceeds in the same order as the plotted individual amplifier values. The crosses show the calculated sums of the individual contributions assuming in-phase addition. Agreement between these values and the measured sums is well within the accuracy of the measurements, considering the difficulties involved and the length of time required to complete the run. The dots show the resultant modulation which would be obtained by adding the power in the individual components instead of the voltages, which would be the expected result for a large number of components with random phase angles. The modulation thus calculated is much smaller than the measured values indicating that a hypothesis of random phasing is untenable for this product.

In actual systems both the magnitude and phase shift of modulation products in the different repeater sections exhibit variations because of non-uniform output levels, differences in tubes and other amplifier parts, and unequal repeater spacings. The addition factor for converting system requirements to single amplifier requirements should therefore contain a marginal allowance for these irregularities in performance.

Summarizing our conclusions on addition of modulation from multiple sources, we may state that the third order requirement invokes the most severe condition—that of in-phase addition. Second order products on the other hand will have enough phase shift, either inherent or from simple reversals of terminals at alternate sections, to make the addition no more rapid than on a power basis. In fact if there is a high degree of similarity with respect to both amplitude and phase increment of products from successive amplifiers throughout the system, the total second order modulation may be much less than calculated from addition of power. In setting the requirements which each amplifier must meet, marginal allowances should be made for differences in lineup throughout the system and aging effects which may take place after the amplifier is put in service.

Let A_x represent the ratio expressed in db between the total x-type modulation received from the system and the contribution of x-type from one amplifier, assuming the amplifiers contribute equally. For example, if there are K amplifiers in the system and if the contributions to the product add in phase $A_x = 20 \log_{10} K$. If power addition occurs, $A_x = 10 \log_{10} K$. A favorable set of phase angles may reduce this factor by an amount depending on the uniformity of the repeaters. If the system is divided into K_1 links having K_2 amplifiers in each link, with phase shifts and changes of frequency allocations of individual channels present at the link junctions such that amplitude addition occurs within links and power addition from link to link, $A_x = 10 \log_{10} K_1 + 20 \log_{10} K_2$. We shall also introduce a lineup factor F_x defined as the number of db by which the x-type product requirement must be increased to allow for irregularities in lineup operating levels of the amplifiers. These may be due to initial differences in repeaters or cable sections and to subsequent changes which may occur because of aging effects. If H_{x0} represents the requirement on the ratio of fundamental to x-type product in the output of a single amplifier when one mw. of test tone power is delivered at zero level,

$$H_{x0} = H_x + A_x + F_x, \tag{8.1}$$

where H_x is the system requirement given by (7.8).

9. Testing Methods

It is difficult to test a broad band carrier system under conditions simulating normal operation because of the large number of independent conversations required to load the channels. We have seen that much information applicable to speech load can be deduced from current analyzer measurements of modulation products when discrete frequencies are applied. Since rather extensive calculations are required to evaluate the performance of the system from single frequency data, an overall test under conditions comparable to actual operation has considerable value as a check. A convenient method of simulating the speech load in the high frequency medium by means of a uniformly distributed spectrum of energy such as thermal noise or the output of a gas tube applied through a narrow band elimination filter has been developed [8] for this purpose. The band elimination filter suppresses the energy which would fall in several adjacent channels; hence anything received in these channels at the receiving end of the line is introduced by the system. We can thus measure interchannel modulation if it exceeds the background of noise from other sources. A

[8] E. Peterson, *Bell Laboratories Record*, Nov. 1939, Vol. 18, No. 3, pp. 81–83.

summation is obtained over all types of products but the predominant order may be distinguished by the power law followed. Effects not simulated are the frequency distribution of speech energy within individual channels and the idle periods which occur in various channels during normal operation, but these effects are minor in a system with a large number of channels. The noise loading method is particularly valuable on a multirepeater line, in which modulation measurements made with discrete frequencies may show large variations with frequency because of phasing between the various sources. Loading with a flat band of energy secures an average of these variations over the frequency range used.

Acknowledgments

Contributions to the solution of the various problems discussed here have been made by many members of the Bell Telephone Laboratories. The author wishes to take this opportunity to acknowledge his indebtedness to those of colleagues who have participated in the development of the ideas here discussed.

Other References

Following is a list of published papers (excluding those to which reference has already been made in the text) relating to various aspects of the general problem discussed here:

F. Strecker, "Nichtlineares Nebensprechen bei der gemeinsamen Übertragung mehrerer modulierten Trägerwellen," *Hoch. u. Elek.*, **49** (1937), 5, pp. 165–171.

H. Jacoby and G. Günther, "Über die Wahrscheinlichkeit der in Trägerfrequenz-Vielfachsystemen auftretenden linearen und nichtlinearen Spannungen," *Hoch. u. Elek.*, **52** (1938), 6, pp. 201–209.

E. Hölzler, "Das Nichtlineare Nebensprechen in Mehrfach-Systemen mit Übertragenen Trägern," *Hoch. u. Elek.*, **52** (1938), 4, pp. 137–142.

H. F. Mayer and D. Thierbach, "Über den Einfluss von Nichtlinearität und Wärmerauschen auf die Reichweite von Trägerfrequenz-Vielfach-Systemen," *E. F. D.*, **48** (1938), pp. 6–12.

H. Tischner, "Zur Berechnung und Messung nichtlinearen Verzerrungen in Trägerfrequenten Übertragungs-systemen," *Zeitschr. f. techn. Physik*, **19** (1938), 11, pp. 425–429.

Yonezawa and Hirayama, "Relation Between the Non-Linear Distortion and the Crosstalk on Multiplex Transmission," *Nippon Elec. Comm. Eng.*, June, 1938.

Yonezawa, "Non-Linear Distortion and Crosstalk Frequency Groups in Multiplex Carrier Transmission," *Nippon Elect. Comm. Eng.*, September, 1938.

Matzumae and Yonezawa, "Equipments for Multiplex Carrier Telephony on Ultra Short Wave," *Nippon Elect. Comm. Eng.*, Feb. 1939.

APPENDIX A

Evaluation of Volume Corresponding to Mean Power When Volume Distribution is Normal

The successive steps in the evaluation of (3.2) are as follows:

$$\overline{\text{antilog}_{10} V/10} = \overline{10^{V/10}} = \overline{\exp\left(\frac{V \log_e 10}{10}\right)}$$

$$= \int_{-\infty}^{\infty} \exp\left(\frac{V \log_e 10}{10}\right) p(V) dV$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{V}{10}\log_e 10 - \left(\frac{V-V_0}{2\sigma^2}\right)^2} dV$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\log_e 10}{10}\left(V_0 + \frac{\sigma^2}{20}\log_e 10\right)} \int_{-\infty}^{\infty} e^{-\frac{\left(V - V_0 - \frac{\sigma^2}{10}\log_e 10\right)^2}{2\sigma^2}} dV$$

$$= e^{\frac{\log_e 10}{10}\left(V_0 + \frac{\sigma^2}{20}\log_e 10\right)} = 10^{\frac{1}{10}\left(V_0 + \frac{\sigma^2}{20}\log_e 10\right)}.$$

Hence

$$V_{0p} = 10 \log_{10} \overline{\text{antilog}_{10} \frac{V}{10}} = V_0 + \frac{\sigma^2}{20} \log_e 10.$$

APPENDIX B

Counting of Modulation Products

Consider an N-channel system with carrier frequencies n_1p, $(n_1 + 1)p$, $\cdots$, n_2p, where $n_2 = n_1 + N - 1$. Let q_n represent the frequency of the signal impressed on the voice frequency channel associated with the carrier frequency np. The wave to be amplified is then

$$e_g = \sum_{n=n_1}^{n_2} Q_n \cos\left[(np + q_n)t + \theta_n\right]. \tag{1}$$

The phase angles are of no consequence if the signal frequencies are incommensurable; hence we shall simplify the notation by setting $\theta_n = 0$. The square of the single series in n may be written as a double series in m and n, thus:

$$a_2e_g^2 = \frac{a_2}{2}\sum_{n=n_1}^{n_2} Q_n^2 + \frac{a_2}{2}\sum_{n=n_1}^{n_2} Q_n^2 \cos(2np + 2q_n)t$$

$$+ a_2 \sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_mQ_n \cos[(m+n)p + q_m + q_n]t$$

$$+ a_2 \sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_mQ_n \cos[(m-n)p + q_m - q_n]t. \quad (2)$$

Similarly the term $a_3e_g^3$ may be written as a triple series in l, m, n as follows:

$$a_3e_g^3 = \frac{3a_3}{4}\sum_{n=n_1}^{n_2} Q_n^3 \cos(np + q_n)t$$

$$+ \frac{a_3}{4}\sum_{n=n_1}^{n_2} Q_n^3 \cos(3np + 3q_n)t$$

$$+ \frac{3a_3}{2}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_m^2Q_n \cos(np + q_n)t$$

$$+ \frac{3a_3}{2}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_mQ_n^2 \cos(mp + q_m)t$$

$$+ \frac{3a_3}{4}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_m^2Q_n \cos[(2m+n)p + 2q_m + q_n]t$$

$$+ \frac{3a_3}{4}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_m^2Q_n \cos[(2m-n)p + 2q_m - q_n]t$$

$$+ \frac{3a_3}{4}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_mQ_n^2 \cos[(m+2n)p + q_m + 2q_n]t$$

$$+ \frac{3a_3}{4}\sum_{m=n_1+1}^{n_2} \sum_{n=n_1}^{m-1} Q_mQ_n^2 \cos[(m-2n)p + q_m - 2q_n]t$$

$$+ \frac{3a_3}{2}\sum_{l=n_1+2}^{n_2} \sum_{m=n_1+1}^{l-1} \sum_{n=n_1}^{m-1} P_lP_mP_n$$
$$\times \{\cos[(l+m+n)p + q_l + q_m + q_n]t$$
$$+ \cos[(l+m-n)p + q_l + q_m - q_n]t$$
$$+ \cos[(l-m+n)p + q_l - q_m + q_n]t$$
$$+ \cos[(l-m-n)p + q_l - q_m - q_n]t\}. \quad (3)$$

It is now a straightforward, though somewhat tedious process, to count the total number of possible products of each type falling in individual channels. The arrangement of terms above is such that $l > m > n$; this forms a convenient way of insuring that no product is counted twice. We shall illustrate by taking a simple case—the second order sum product, or $(A + B)$-type. Let kp represent the carrier frequency of the channel in which we wish to determine the number of possible $(A + B)$-type products. This means that in (2) we wish to find the number of terms in the third summation in which $m = n = k$. The resulting sum becomes:

$$\sum_{m=n_1+1}^{n_2} (\cdot \quad \leq k - m \leq m - 1). \tag{4}$$

That is, there are as many terms as there are integer values of n satisfying the simultaneous inequalities,

$$\begin{bmatrix} n_1 + 1 \leq m \leq n_2 \\ \dfrac{k+1}{2} \leq m \leq k - n_1 \end{bmatrix} \tag{5}$$

The number of terms is zero if $k > 2n_2 - 1$, because the lower limit of the second inequality exceeds the upper limit of the first. If $n_1 + n_2 \leq k \leq 2n_2 - 1$, the upper limit of the first inequality and the lower limit of the second inequality are governing, and the number of terms is $n_2 - I(k/2)$, where $I(x)$ is a symbolic representation for the largest integer $\leq x$. If $2n_1 \leq k \leq n_1 + n_2$, the second inequality is governing and the number of terms is $I\left(\dfrac{k+1}{2}\right) - n_1$. If $k \leq 2n_1$, the number of terms is zero.

In a similar manner the more complicated sums representing the number of third order products can be evaluated. It is to be noted that contributions to a particular type can come from more than one of the sums listed. For example, the $(A + B - C)$-type is made up of the summations from the $l + m - n$, $l - m + n$, and $l - m - n$ terms. In fact all these are of $(A + B - C)$-type except those in which $l - m - n$ is negative. The latter, since only positive values of frequency are significant, are of $(A - B - C)$-type. An $(A - B - C)$-type product differs from $(A + B - C)$-type not only in S.T.M.F., but also in manner of addition of contributions from a multi-repeater line as discussed in Section 8.

As an alternative to an actual count of the products falling in a channel, it is possible to approximate the sum by an integration

process when the number of channels is large. This is an especially valuable simplification for products of high order for which the counting becomes very tedious. Suppose there are D components in unit band width in the range a to b. Let $x_1, x_2, \cdots x_n$ be n frequencies such that $a \leq x_1 < x_2 < \cdots < x_n \leq b$. The number of products of form $m_1x_1 + m_2x_2 + \cdots + m_nx_n$ which can be formed by selecting fundamentals from the n bands of width dx_1 at x_1, dx_2 at x_2, $\cdots$ dx_n at x_n is $D^n dx_1 dx_2 \cdots dx_n$. To count the total number of such products which can be formed in the band a to be such that the resultant frequency lies in the band x_0 to x, we form the integral

$$g(x, x_0) = D^n \int_a^b dx_n \int_a^{x_n} dx_{n-1} \cdots \int_a^{x_2} \varphi(x, x_0, x_1, x_2 \cdots x_n) dx_1, \quad (6)$$

where

$$\varphi(x, x_0, x_1, x_2 \cdots x_n) = \begin{pmatrix} 1, x_0 \leq m_1x_1 + m_2x_2 + \cdots + m_nx_n \leq x \\ 0, \text{ otherwise} \end{pmatrix}. \quad (7)$$

A suitable representation of φ is furnished by

$$\varphi(x, x_0, x_1, x_2 \cdots x_n) = \frac{1}{2\pi i} \int_C \left(e^{-ix_0z} - e^{ixz} \right) e^{iz \sum_{r=1}^{n} m_r x_r} \frac{dz}{z}, \quad (8)$$

where C is a contour going from $z = -\infty$ to $z = +\infty$ and coinciding with the real axis except for a downward indentation at the origin. To obtain the number of products $\nu(x)dx$ falling in band of width dx at x, we write

$$\nu(x) = \operatorname*{Lim}_{\Delta x = 0} \frac{g(x, x - \Delta x)}{\Delta x}$$

$$= \frac{D^n}{2\pi} \int_C e^{-ixz} dz \int_a^b dx_n \int_a^{x_n} dx_{n-1} \cdots \int_a^{x_2} e^{iz \sum_{r=1}^{n} m_r x_r} dx_1. \quad (9)$$

If the spacing between carrier frequencies is used as the unit band width, we may set $D = 1$, $a = n_1$, $b = n_2$, $x = k$, in (9) and obtain the limiting forms of the results given in Table II as the number of channels is made large. Evaluation of (9) is easily carried out by means of the relation:

$$\int_C \frac{e^{ixz} dz}{z^m} = \begin{bmatrix} \dfrac{2\pi i^m x^{m-1}}{(m-1)!}, & x \geq 0 \\ 0, & x < 0 \end{bmatrix}. \quad (10)$$

As an example, suppose it is desired to calculate the approximate number of $(2A + B)$-type products falling in channel number k when the number of channels is large. If $x_2 > x_1$, this type of product may be either of form $2x_1 + x_2$ or $2x_2 + x_1$. Both are included by the expression

$$
\begin{aligned}
\nu_{2A+B}(k) &= \frac{1}{2\pi}\int_C e^{-kz}dz \int_{n_1}^{n_2} dx_2 \int_{n_1}^{x_2} \left[e^{iz(2x_1+x_2)} + e^{iz(x_1+2x_2)} \right] dx_1 \\
&= \frac{D^n}{4\pi}\int_C \frac{e^{-ikz}}{i^2 z^2}\left[e^{2in_2 z} - e^{i(2n_1+n_2)z} - e^{i(n_1+2n_2)z} + e^{3in_1 z} \right] dz \\
&= \begin{cases} 0, \; k < 3n_1 \\ \tfrac{1}{2}(k - 3n_1), \quad 3n_1 < k < 2n_1 + n_2 \\ \tfrac{1}{2}(n_2 - n_1), \quad 2n_1 + n_2 < k < n_1 + 2n_2 \\ \tfrac{1}{3}(3n_2 - k), \quad n_1 + 2n_2 < k < 3n_2 \\ 0, \quad k > 3n_2. \end{cases} \qquad (11)
\end{aligned}
$$

The method may be generalized to include the calculation of the modulation spectra produced by fundamentals having specified arbitrary spectra by inserting the appropriate function of the power in unit band width at $x_1, x_2, \cdots x_n$ in the integrand of (6).

Radio Extension Links to the Telephone System

By R. A. HEISING

TO the general public, the word *radio* means broadcasting, and a *radio set* means a radio receiver for listening thereto. The average man never has any direct contact with a radio transmitter, nor with the radio telegraph which preceded the radio telephone and so utilizes the all inclusive word "radio" for that one part of it which he sees, buys, and uses.

The radio engineer is not quite in that class. There is, however, much in radio with which he is unacquainted. The radio field has become so broad and extensive that it is physically impossble for anyone to keep abreast of the whole art. Since the application of radio to telephone links is a specialized field, undoubtedly much of its history and technical developments is known only sketchily or not at all to many engineers. This paper, therefore, is planned to cover this field briefly, show its general development, and describe in principle a number of devices developed for use therein, most of which are seldom if ever used in the field of broadcasting.

The radio telephone was not one of those devices that an inventor springs upon an unexpecting world. On the contrary it was expected, and was the object of search and investigation for years before a practical form appeared. Because the wire telephone followed the wire telegraph, technical men expected the radio telephone to follow the radio telegraph as soon as the latter had been practically demonstrated. Telephone men developed an interest in it as soon as it was suggested. Telephony over large bodies of water, over difficult terrain, and to moving conveyances was difficult or impossible for wire telephony, and the telephone man was intrigued by the possibility of providing his circuits without the use of conducting wires.

During the first few years of this century, several radio telephone systems were technically demonstrated but were found impractical. In 1912 an important step occurred. The audion, invented by DeForest, was brought to the attention of the American Telephone and Telegraph Company. It appeared to have possibilities making it superior to the mechanical and the arc repeaters for wire telephone lines. A telephone repeater, or amplifier, was a main object of search at that time by telephone men. The audion became the subject of study in the Research Department of the American Telephone and

Telegraph Company and Western Electric Company immediately, and within a year and a half went through some rapid transformations. A high vacuum and increased electron emission were provided by H. D. Arnold and A. M. Nicolson, while a practical circuital theory was provided by H. J. Van der Bijl. The internal arrangement was engineered and a socket and base developed. This improved vacuum tube was put into use on the commercial telephone lines in the latter half of 1913 as a telephone amplifier and was the first commercial use of the high vacuum tube. This vacuum tube amplifier contributed to the establishment of the original transcontinental wire telephone line which carried its first messages in July 1914.

The improved vacuum tube, during its period of development, appeared to have possibilities as a generator of sustained oscillations and suggested to telephone engineers that it might be much more useful in radio than it had been up to that time. With this in mind the A. T. & T. Company decided to start work in that direction and as one result a number of new engineers, including the writer, were employed and began work in the Research Department of the Western Electric Company in the middle of 1914. Developments on the radio telephone moved rapidly. Early in 1915 plans were made and active work was started for field trials. A transmitting station was established at Montauk, L. I., and a receiver located on Hotel DuPont in Wilmington, Delaware. On April 4, 1915, speech was transmitted from Montauk to Wilmington, a distance of 220 miles. Connections were made with telephone lines at both ends to show its possibilities as a link in a telephone circuit.

There followed tests to Jekyl Island, off the coast of Georgia, about 800 miles, and then work was started for a transoceanic test. To transmit across the ocean required more power and a larger antenna. In order to avoid the antenna expense, arrangements were made with the Navy Department to use the Arlington antenna for transmitting, and to use Naval radio stations at San Francisco, San Diego, Panama and Honolulu for receiving locations. Observers with radio receivers were dispatched to these four receiving locations while a fifth expedition was sent to Paris where in spite of the war the French Government kindly allowed listening on the Eiffel Tower antenna. At Arlington the Western Electric Research Department (now part of Bell Telephone Laboratories, Inc.) installed a vacuum tube transmitter, and proceeded to make one-way tests. In August 1915 speech was understood at Panama, and in September a one-way demonstration was made across the continent, receiving at San Francisco. Within a few days speech was heard in Honolulu and then in Paris. The tests

showed that transoceanic telephony was possible and indicated some of the difficulties that had to be overcome.

The radio transmitter in these tests deserves a few words because of its novelty and because in one respect it has never been equaled. The carrier was modulated at a relatively low level and then amplified. The final stage of amplification contained 550 tubes in parallel which in number appears to be an all time record. Each tube was capable of delivering only 15 or 20 watts peak h.f. power which would give a power rating on a telephone basis of about $2\frac{1}{2}$ kw.

With these tests completed, transoceanic telephony withdrew into the laboratory for almost eight years while further intensive work was carried out. The second step in public occurred in January 1923

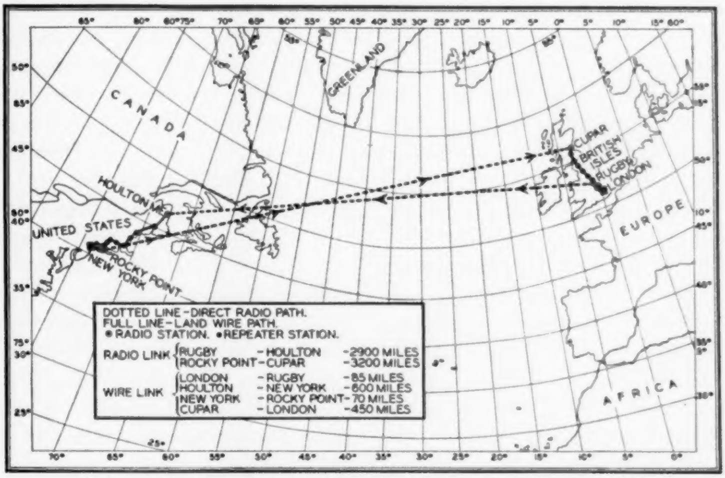


Fig. 1—The first transoceanic radio telephone circuit opened for commercial use January 7, 1927.

when a second transoceanic test was made. A 200 kw. single sideband transmitter had been constructed and installed at Rocky Point, Long Island, while engineers with receiving equipment journeyed to London. A demonstration was given to government engineers and to newspaper reporters over there to show that practical transoceanic telephony was possible and to interest them in constructing a return circuit. The British government was interested and with our assistance took up the matter of providing a transmitter and receiver for their end. Three years were required for this third step and in February 1926 the first two-way radio telephone conversations were held between the United States and England. Commercial service opened in January 1927. See Fig. 1.

With the first transoceanic circuit established, further circuits followed rapidly. For a number of years prior to 1927 investigations had shown that short waves could travel enormous distances with very much less attenuation than the long waves. Telephone engineers conducted a series of long distance tests which laid the foundation for developing circuits on these short waves with much less power. As a result a short wave transoceanic telephone circuit, the first of its kind, was opened on June 6, 1928, between the United States and England. The Germans followed by establishing a circuit to Buenos Aires in December of that year. The Dutch established one to Bandoeng in January of the following year, 1929. Then another circuit was opened up between the United States and London in June of 1929. The circuit from Madrid to Buenos Aires was established that same year, and there followed very rapidly circuits from London to various British Colonies, a circuit from New York to Bermuda, and one from San Francisco to Honolulu, so that as of Jan. 1, 1939 the world was covered by a multitude of circuits as indicated in Fig. 2.* However, the radio circuits of greatest interest to us are those circuits extending from this continent to other continents. These appear in Fig. 2. There are several channels to London, one to Paris (temporarily suspended), one to Rome, one to Australia (temporarily suspended), one to Berlin, one to Switzerland, two channels to Honolulu and a number of circuits to South and Central American countries, circuits to Manila, Bandoeng, Tokyo and Shanghai. There is a circuit from Montreal to London operated by the Canadian Marconi Company and British Post Office. The facilities are now such that from almost any telephone in the United States it is physically possible to talk to almost any telephone in the rest of the world, although due to censorship some of the circuits are not actually in use.

Another use for radio as a link in telephone service is for providing service where wire circuits would be unusually expensive and difficult to maintain. Such a circuit indicated in Fig. 2 runs from Seattle to Juneau, Alaska. It is operated by the Signal Corps and connects with the general telephone system at Seattle. Another circuit is shown in Fig. 3 which runs from Green Harbor to Provincetown, Massachusetts, a distance of 24 miles across Cape Cod Bay. This circuit supplements the wire lines which reach Provincetown by a roundabout path by land around the south side of the Bay. The radio link provides a very desirable alternate route to points on the Cape and has been useful in a number of emergencies in maintaining

* Explanation.—On account of unsettled conditions, has not been brought up-to-date.

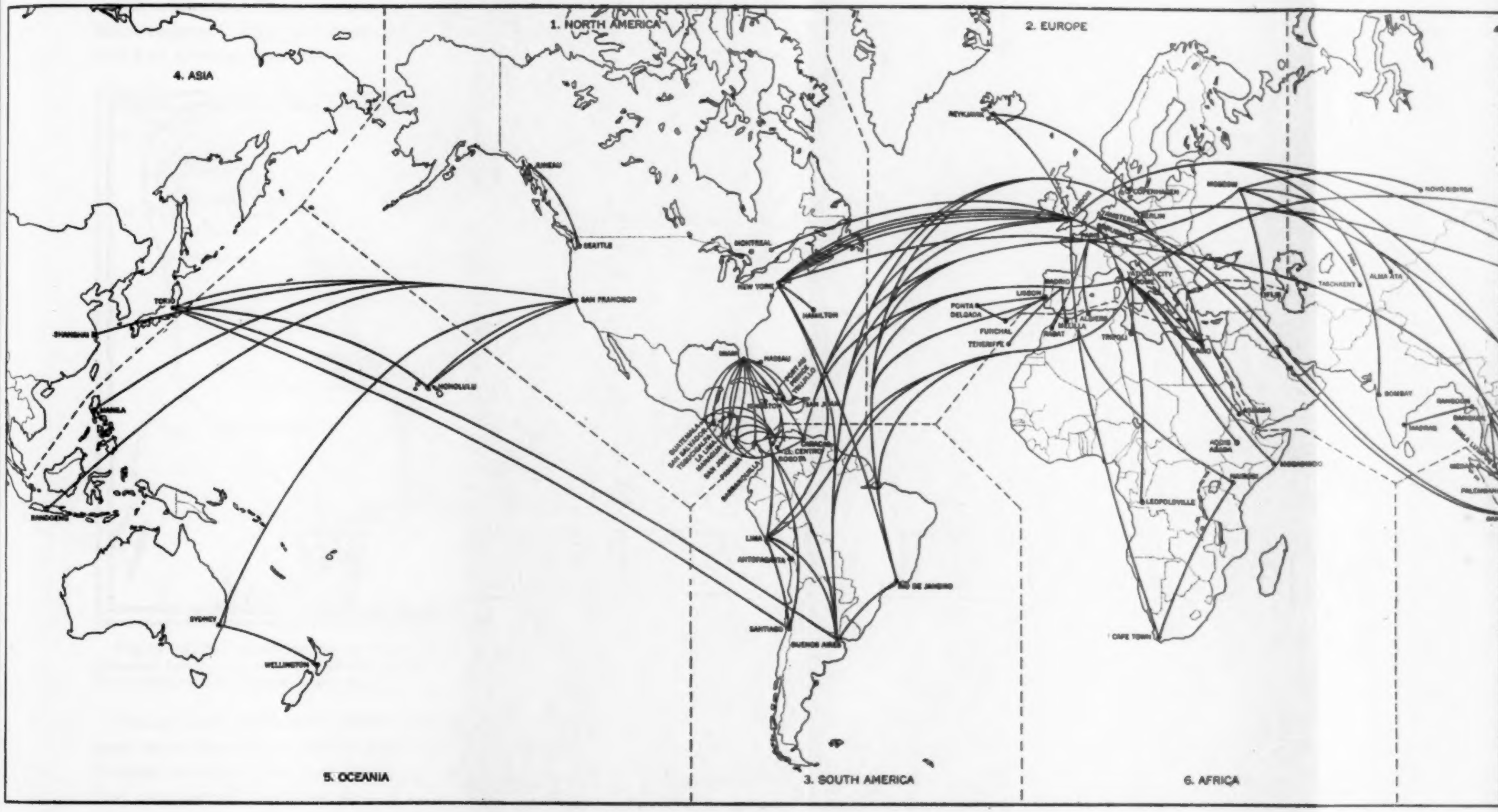


Fig. 2—Transoceanic radio telephone circuits in operation January 1, 1939.

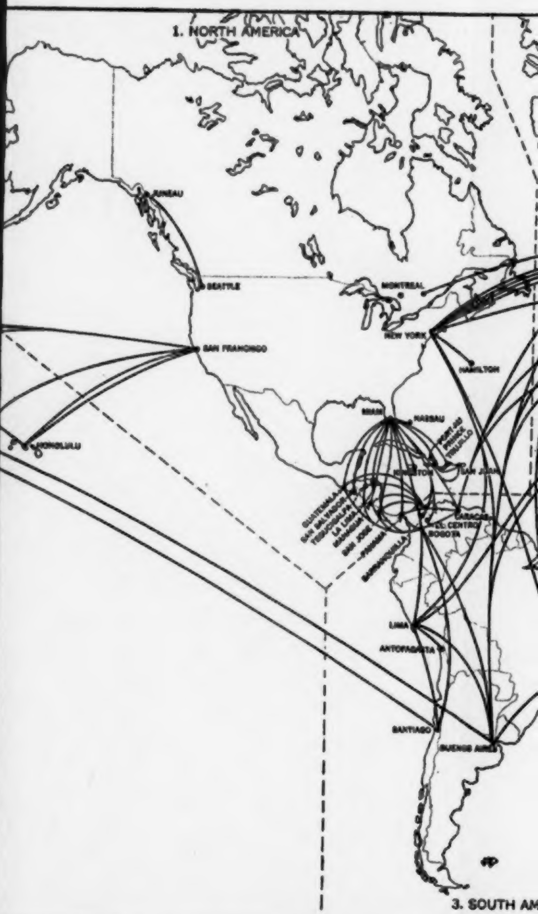
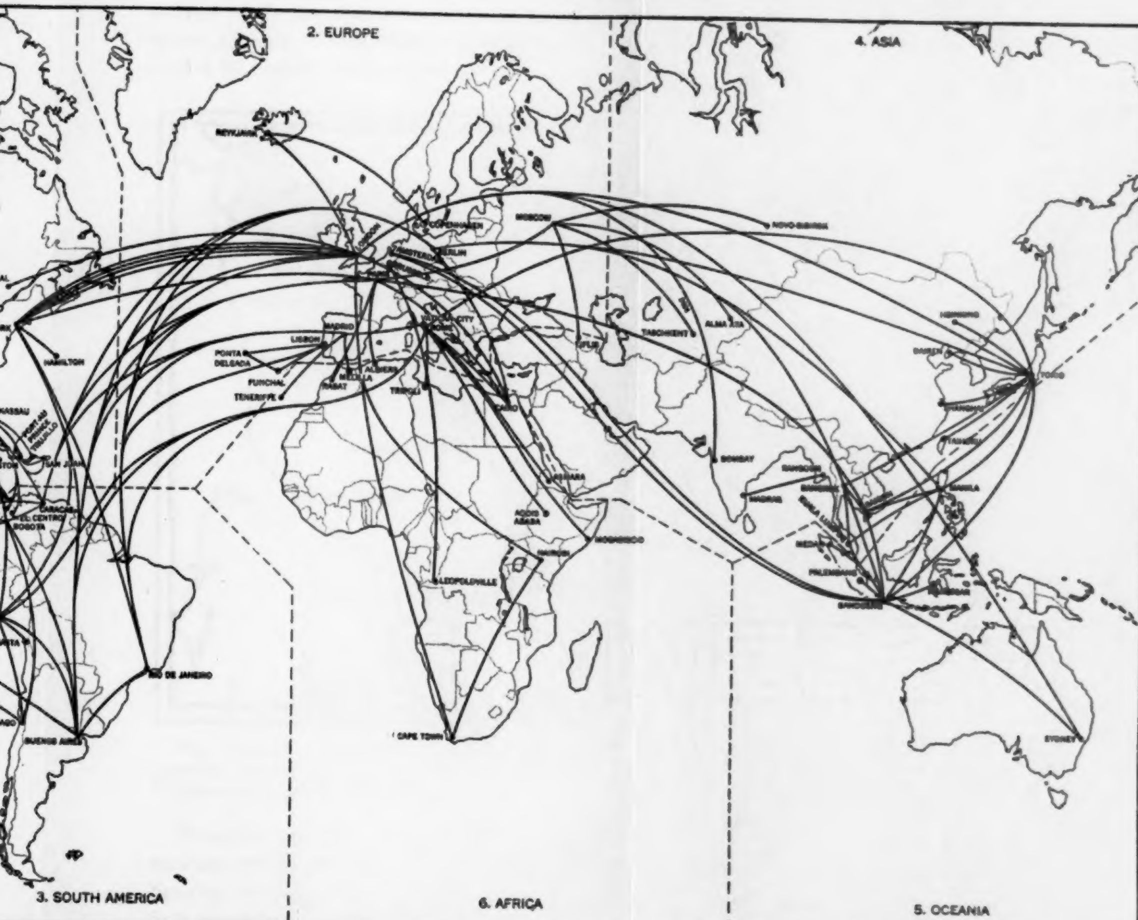


Fig. 2—Transoceanic radio telephone



adio telephone circuits in operation January 1, 1939.

service. During the hurricane of 1938 it provided the only route to Cape Cod for a time. The Provincetown radio link is different from any of the transoceanic links mentioned previously in that it operates in a third region of the radio spectrum known as the ultra-short-wave region while the transoceanic circuits are in the short-wave and long-wave regions. This circuit operates on 63 and 65 megacycles, 4.75 and 4.61 meters, respectively.

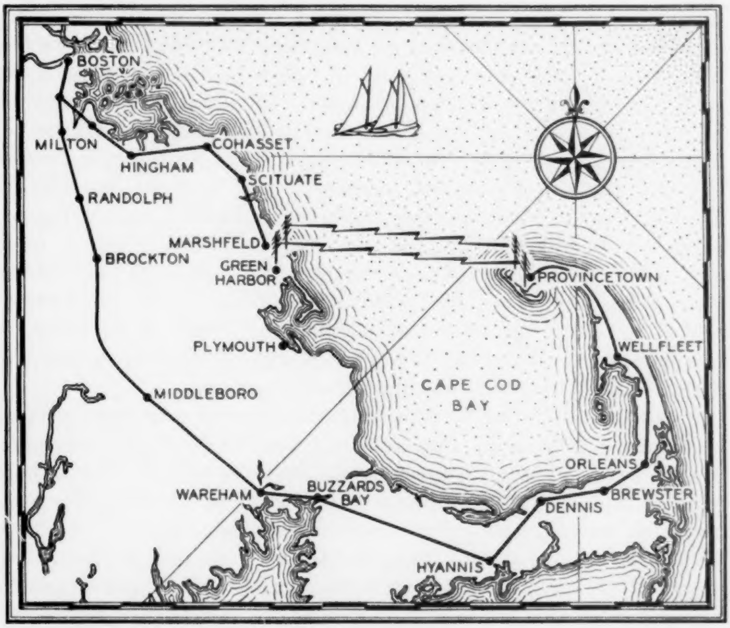


Fig. 3—Radio telephone circuit from Provincetown to Green Harbor, Massachusetts, connecting with the telephone line to Boston. Land wire route between Provincetown and Boston is also shown.

Transoceanic and other point-to-point services were not the only services envisaged by radio engineers prior to the era of broadcasting. Service to ships was considered an important use for radio. The first commercial telephone service to ships was the service established with transatlantic liners in December of 1929. Most of the larger transoceanic liners are now equipped for radio telephone communication with both shores. A few years after this service was established, a ship radio telephone service of a more local character was initiated to serve fishing fleets off the New England coast. The radio station

for this service was established near Boston. The necessary equipment for boats was also developed. The fishing boats by means of this service can keep in touch with the fish markets and can take advantage of rises in prices. They also find it convenient for communicating with each other when schools of fish are found and that has also been a help in their operations. There have been a number of occasions also in which it has resulted in the saving of lives at sea, as the radio was used to notify the shore station in case of accident and the shore station called other vessels and sent them to the rescue. This service to fishing vessels was then extended to other coastwise vessels, yachts, tugs, etc., so that there has gradually developed an extensive radio service of this type on both of our coasts. Radio stations are located not only in Boston now but as indicated in Fig. 4 there are stations at New York, Ocean Gate, N. J., Wilmington, Del., Norfolk, Charleston, S. C., Miami, New Orleans, Galveston, Los Angeles (San Pedro), San Francisco and Seattle. Stations are under construction at Tampa, Fla., Astoria and Portland, Oregon. Service is now given to more than 2,000 vessels, there being 200 tugs, 1,100 yachts, 100 steamships, 400 fishing vessels and numerous others, police boats, pilot boats, barges, launches, etc. The largest number of vessels so equipped for communication with shore are grouped around New York and San Pedro, there being about 600 in each of these areas.

In this type of service each shore transmitter and each shore receiver is assigned a frequency. Any ship may provide itself with frequency control crystal elements for communicating with as many of the shore stations as it desires. Coastwise vessels in traveling along the coast may thus keep in touch with their nearest shore station. In New York there are two such circuits provided with transmitters located on Staten Island, and there are receivers located at four places around the harbor for each of the circuits so that the low-powered ship transmitters may reach the nearest receiver while the higher-powered shore transmitters reach the ship receivers directly.

On the United States side of the Great Lakes, connecting telephone companies operate coastal harbor radio telephone stations at Lorain, Ohio, Duluth, Port Washington, Wis., Lake Bluff, Ill. and Mackinac Island.

The use of radio in the telephone system brings forth a number of problems. First of all, to provide a radio circuit for a telephone conversation there are required two radio transmitters and two radio receivers. The transmitters and receivers must be so located and so designed and operated that one person at one end of the circuit may speak to a second person at the other end and the second one to the

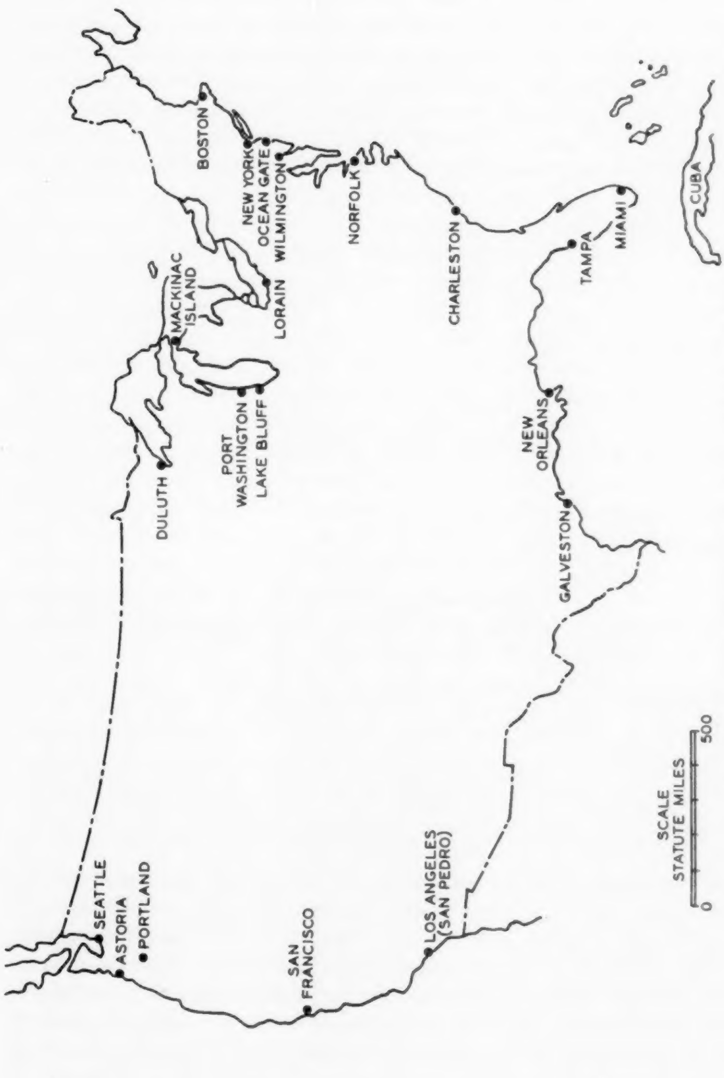


Fig. 4—Coastwise boat radio service, and other radio links.

first even though either or both of the radio equipments be entirely outside of the manual control of the speakers and be many miles away. The problems involved have given rise to the development of many pieces of apparatus which are seldom used in the broadcasting field. It is the intention, therefore, in this paper to review some of these devices and tell briefly why they are used and what they do.

To begin with, attention is called to three diagrams in Fig. 5. These three diagrams indicate three of the many ways in which a transmitter

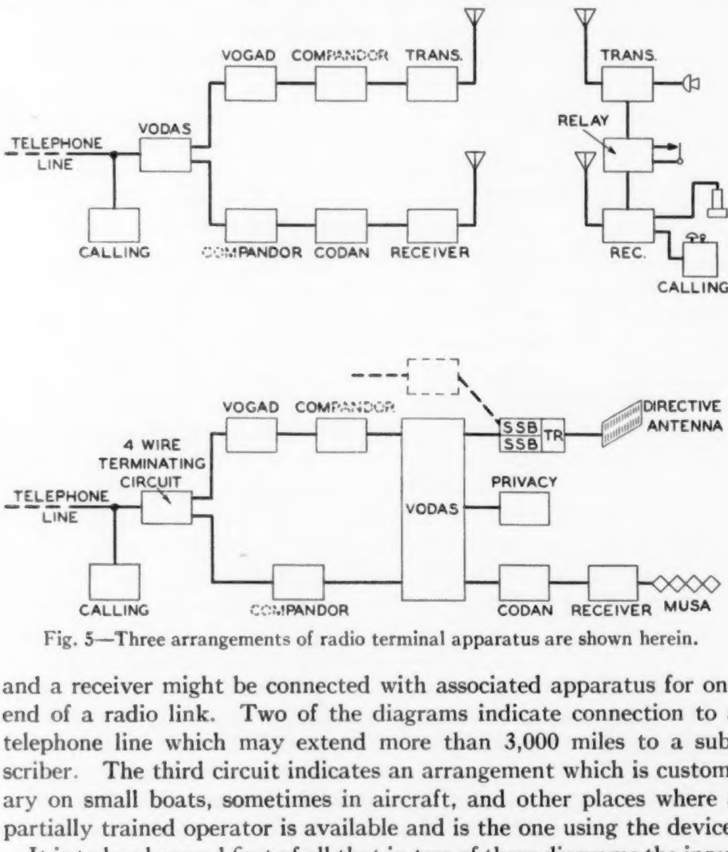


Fig. 5—Three arrangements of radio terminal apparatus are shown herein.

and a receiver might be connected with associated apparatus for one end of a radio link. Two of the diagrams indicate connection to a telephone line which may extend more than 3,000 miles to a subscriber. The third circuit indicates an arrangement which is customary on small boats, sometimes in aircraft, and other places where a partially trained operator is available and is the one using the device.

It is to be observed first of all that in two of these diagrams the input to the transmitter and the output of the receiver are connected to the same telephone line. This necessary connection leads to difficulties. Signals from a subscriber on the telephone line operate the

radio transmitter. Some of the radiated energy from the transmitter impinges on the receiver. If the receiver should be tuned to the same wave-length the signals will then get back onto the telephone line and some will again go to the transmitter, producing by this circular path a singing circuit. A circuit so constructed will be entirely useless due to the singing produced. Now it appears possible to use a hybrid coil to connect a balancing circuit to the telephone line, with conjugate connections to the transmitter and to the receiver so that the incoming signals from the receiver will not go to the transmitter. Such a hybrid circuit will work provided it can be balanced and maintained in balance. However, anyone who has tried balancing such circuits knows that it is generally not practicable to provide a balanced circuit suitable for all wire line connections and for the variable gains in the radio link. Additional means must therefore be used. Now, of course, it is possible to operate the incoming radio circuit and, therefore, the receiver on a different frequency as is usually done, in which case the signals from the local transmitter will be tuned out. However, if a similar system is used at the other end of the radio link the signals from the near end transmitter will come in on the far end receiver, will again go out on the far end transmitter, will come back into the near end receiver whence they get back into the near end transmitter, thereby making a loop circuit again which will produce singing even though the round trip path of such a circuit may be 6,000 miles. It is therefore found necessary, when connecting with telephone lines, to provide a system which will at all times keep the incoming energy of the receiver from going out on one's own transmitter.

To accomplish the foregoing is the function of the "Vodas" [1] as indicated on the diagram, a device which connects the telephone line to either the transmitter or the receiver but not to both simultaneously. It must, however, connect them at proper and suitable times so that a two-way conversation can take place. A simple system comes immediately to mind to accomplish this purpose. It is that of a voice-operated relay which throws the telephone line from the receiver to the transmitter whenever the speaker on that end speaks, with the relay making the reverse connection when he stops speaking. Such a simple circuit has been used in some cases but has been found not to be adequate for general use. To begin with, the line is not switched until part of what the speaker has said has arrived to actuate the relay. Some clipping, therefore, occurs. To make things worse many words begin with sounds of small energy like f's and s's, which may not be sufficient to actuate the relay. The relay will then not operate until the vowel sound following arrives and when the relay does operate the

entire preceding consonant is clipped off. Such clipping is sometimes disconcerting and may impair the intelligibility of the transmitted speech. If the relay is made sensitive enough to operate on the f's and s's another difficulty arises. Relatively low values of room noise, noise induced on wire lines, or the speaker breathing into the microphone may produce enough energy to actuate the relay during listening periods, thus interrupting the conversation. In other words, if the relay is sensitive only to the loud sounds clipping will occur, while if sensitive to the weaker sounds it may be actuated by noise. These difficulties are surmounted by using the more complex circuit termed the "Vodas" as indicated in Fig. 6.

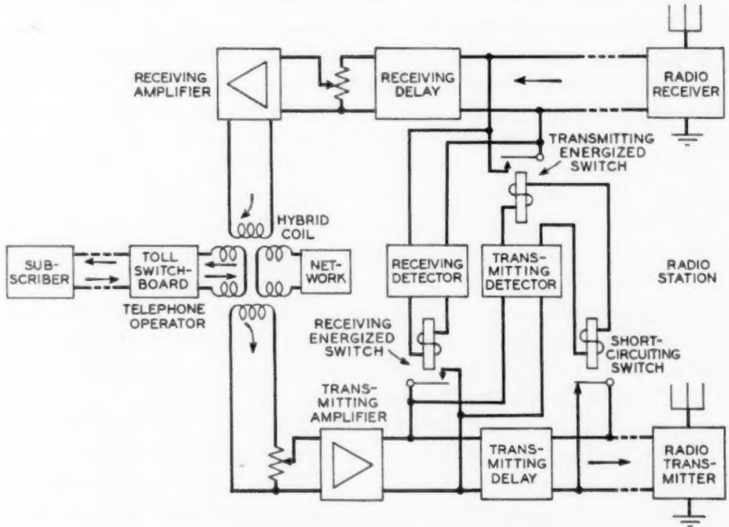


Fig. 6—Vodas (voice operated device anti-singing).

In this diagram the transmitter is located on the bottom branch in which there is interposed a transmitting delay circuit. The incoming speech after passing through the transmitting amplifier operates relays by means of the transmitting detector. The relay is so adjusted as to be operated by the louder sounds in the voice but not by noise. While it is operating upon the vowel sound following a consonant, the consonant may be on its way through the delay circuit so that the relay which normally shuts off signals to the transmitter will actually clear the path to the transmitter in time for the first sounds of a word in most cases. The utilization of the delay circuit can therefore practi-

cally eliminate the clipping and allow of relays being adjusted so as not to be operated by small noises from the telephone line.

This circuit also includes arrangements to prevent other difficulties. It will be observed there are two sets of relays, one operated by the transmitting branch rectifier and one by a receiving branch rectifier. These are so arranged that when speech signals operate the transmitting branch rectifier the receiving line is short-circuited to prevent any signals, such as noise, from the radio receiver reaching the talker or from going out on the transmitting branch to interfere with the transmitted speech. When no talking is occurring at this end of the circuit signals coming in on the radio receiver operate a receiving relay which short-circuits the transmitter circuit so that the received speech will not be retransmitted by the transmitter and so set up a singing condition. This particular diagram indicates the hybrid coil and balancing network which are used to assist in operation but not to provide the main means for preventing the received speech from reaching the transmitter.

This circuit as indicated is about as simple as a satisfactory circuit can be made. Figure 7 indicates a more complex circuit which has a number of advantages, among which is that of connecting in privacy equipment. This circuit allows of using one piece of privacy equipment which is used in the transmitter branch for outgoing signals and is switched to the receiver branch for incoming signals.

Returning now to Fig. 5, attention is called to another device in the first diagram labeled "Vogad."[2] This word comes from the initial letters of the words "voice operated gain adjusting device."[2] This is a type of device which is very useful in telephone practice but is seldom, if ever, used with a broadcasting transmitter. Every telephone user is cognizant of the fact that different people with whom he speaks over the telephone use different intensities of voice; also different lengths of telephone line introduce different amounts of attenuation. If the incoming telephone signals are to operate the radio transmitter to its full modulation capacity some means must be provided to equalize these signals of various levels. It is therefore desirable to have a device which will maintain the output level nearly constant regardless of the variation in the input level due to different speakers and different lengths of telephone lines. This operation is provided by the Vogad indicated in Fig. 8. The channel across the top is the direct path of the speech signals. Within the dotted lines marked "vario repeater" are some elements including the amplifier whose gain is varied to make up for variation in intensities at the input. It is not sufficient to construct an amplifier which will give a large gain

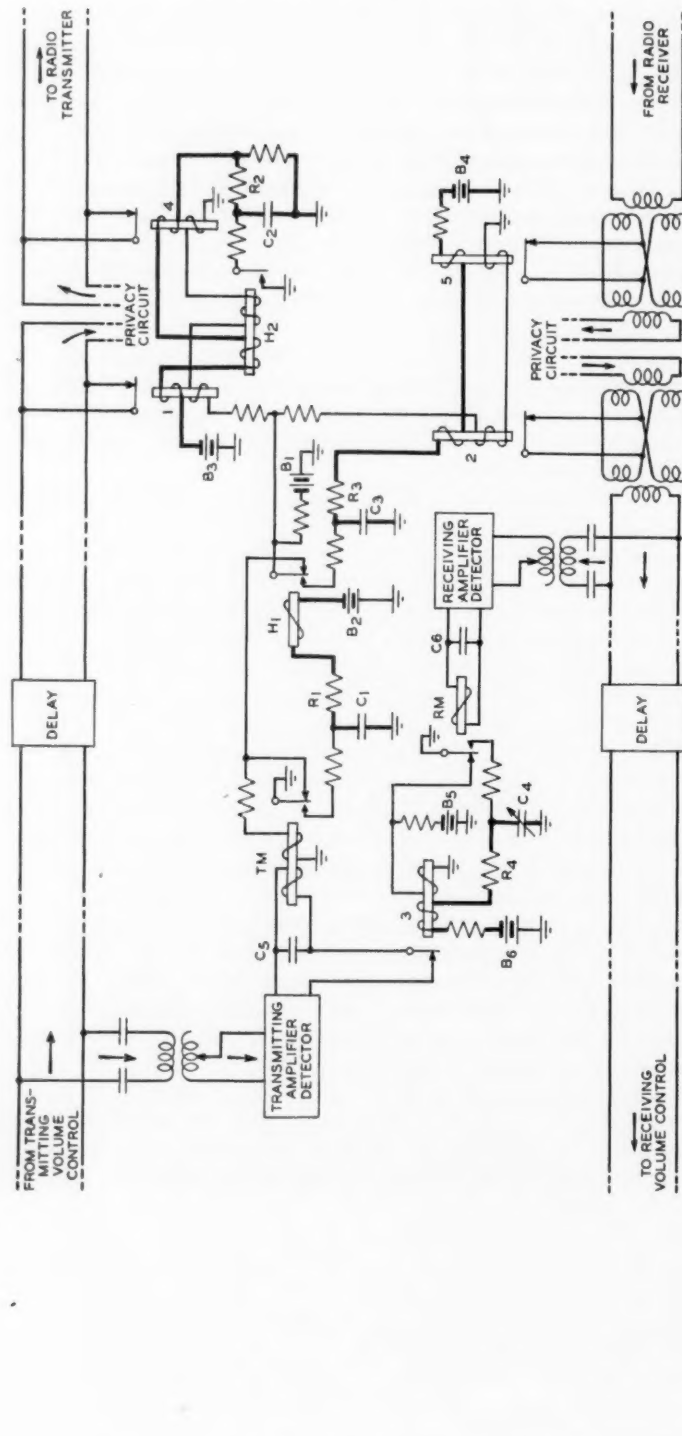


Fig. 7—Improved vodas circuit.

with a small input and small gain with a large input as the use of as simple a circuit as this will cause the amplifier to adjust itself for maximum gain when no input signal is coming in and in that case any noise on the line may be amplified sufficiently to be troublesome. Also, with such an amplifier the gain will become larger whenever the speaker stops or hesitates and will momentarily overload the transmitter with the first syllable on resumption which may result in dis-

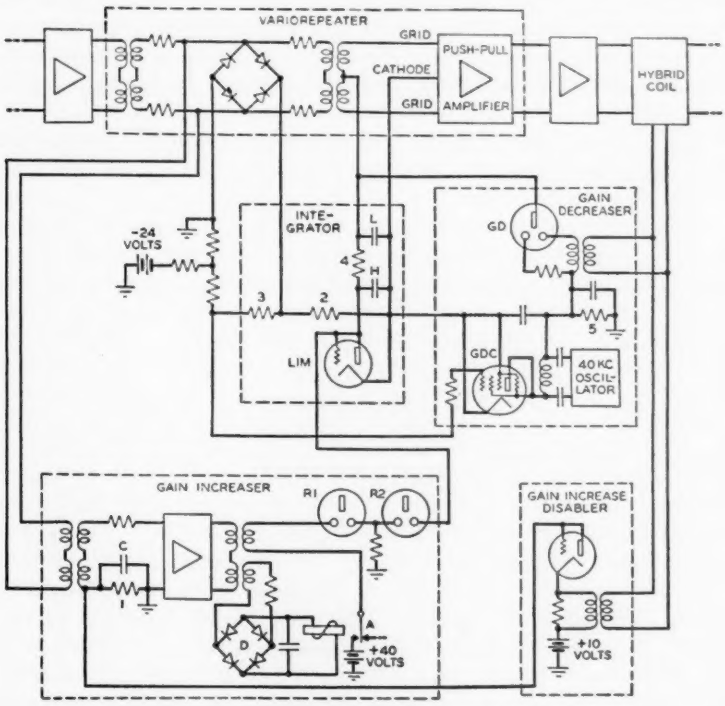


Fig. 8—Vogad (voice operated gain adjusting device).

tortion of considerable consequence. It is therefore desired that this circuit have a maximum gain which will not make noise troublesome and also be so constructed that the gain during any conversation will remain constant even during hesitation and listening periods, and in which the gain will increase only when speech signals become weaker or the gain decrease when speech signals becoming stronger. To accomplish this one element of the circuit marked "Gain Increaser" receives energy from the input and operates through the gas tubes

R1 and R2 causing the gain integrator to increase the gain of the main amplifier when the speech input levels decrease. However, this increase in gain must continue only until sufficient speech volume is going out to modulate the transmitter in an approximately satisfactory manner. At this instant another part of the circuit called the "Gain Increaser Disabler" operated by the output signal comes into play and disables the gain increaser. If the input signal and, therefore, the output signal become louder, then a fourth element, called the "Gain Decreaser," comes into play and begins to reduce the gain of the amplifier. The combination of these control circuits with the main amplifier therefore causes the volume of output signal to be reasonably constant with wide variations in the volume of the input signal and at the same time to hold the gain substantially constant as long as speech signals of the same volume are coming or while no speech signals are coming.

Returning to Fig. 5, note a situation which may be troublesome. Anyone who has operated the earliest broadcast receivers with automatic gain control will remember that when the incoming signal became weak or disappeared the gain of the receiver climbed to such a point as to produce disconcerting noise in the loud speaker. The receivers in all of the systems indicated in this figure contain automatic volume control and if the transmitter from the remote station stops momentarily or if the signals fade out this automatic volume control boosts the gain to such an extent that noise is delivered by the receiver to the telephone line. Such noise may be sufficient to operate the Vodas and in doing so will seize control of the circuit and not allow the signals from the subscriber at this end to reach the transmitter. This would lock up the circuit and put an end to the conversation.

Such a contingency is avoided by the use of the device indicated in the block marked "Codan."[3] The word Codan comes from the initials of the words "carrier operated device anti-noise." The Codan is a device which is operated by the carrier picked up by the receiver and connects the receiver to the telephone line only while a carrier is present. Under these conditions the volume control will go up and down as the carrier goes down and up but if the carrier disappears the Codan disconnects the receiver so that noise will not operate the Vodas and prevent the speech from the subscriber at the near end from being transmitted.

Specifically, circuits operating with ships at sea must have the Codan or its equivalent because the ships usually employ a system in which the carrier is cut off when the ship stops speaking so that the disappearance of the carrier at the receiving station on shore is the

period during which the shore subscriber is expected to talk. The noise at this time in the receiver on shore must not actuate the Vodas or the transmitting branch will be interrupted. A suitable rectifier with a relay can, under certain conditions, be a satisfactory Codan but not in all cases. There is always a certain amount of static, strays and so forth, reaching an antenna and if the relay is adjusted to operate on a very weak carrier the strays and the static may also operate the relay. Now, of course, it is possible to adjust the relay so it will not operate on the noise occurring at a particular time but will operate on a carrier which is slightly stronger. If that adjustment is made during the day on short waves when the noise is low, when night comes the noise level rises and the noise may then be able to operate the relay. It would be necessary with a simple Codan of this type to have the operator continually adjust and readjust the Codan for different parts of the day and night.

However, it is impracticable for an operator to be continuously on watch and continuously and satisfactorily adjust the sensitivity of such a relay, with the consequence that a satisfactory Codan necessarily involves automatic adjustment as provided in the circuit of Fig. 9. In this diagram the part above the middle dividing line is the receiver while the part below is the Codan. The Codan here consists of two parts. It consists of a part which selects the carrier by a crystal filter for operating the relay, and instead of a spring to hold back the armature an electrical arrangement is provided whereby the noise coming through the second part of the circuit will produce current in the relay in the opposite direction. The noise is picked out by another crystal filter Y4 which selects all the energy in the two sideband positions minus the carrier position. This noise is amplified and rectified so that whenever the noise is high it will require a large carrier coming through filter Y3 to operate the Codan relay S2 and when the noise is low through the noise branch a smaller carrier coming through the carrier branch can operate the Codan relay. This Codan therefore automatically adjusts itself to the noise level in the ether so that the carrier can connect the receiver to the telephone line whenever the carrier appears.

Since the development of a successful Codan it has been found practical to dispense with the Vodas at terminals that connect with radio stations which radiate their carrier during transmitting periods only. This is brought about by using the Codan to operate the relay that switches from transmitting to receiving. Since the Codan is operated by the incoming carrier and not by outgoing voice signals,

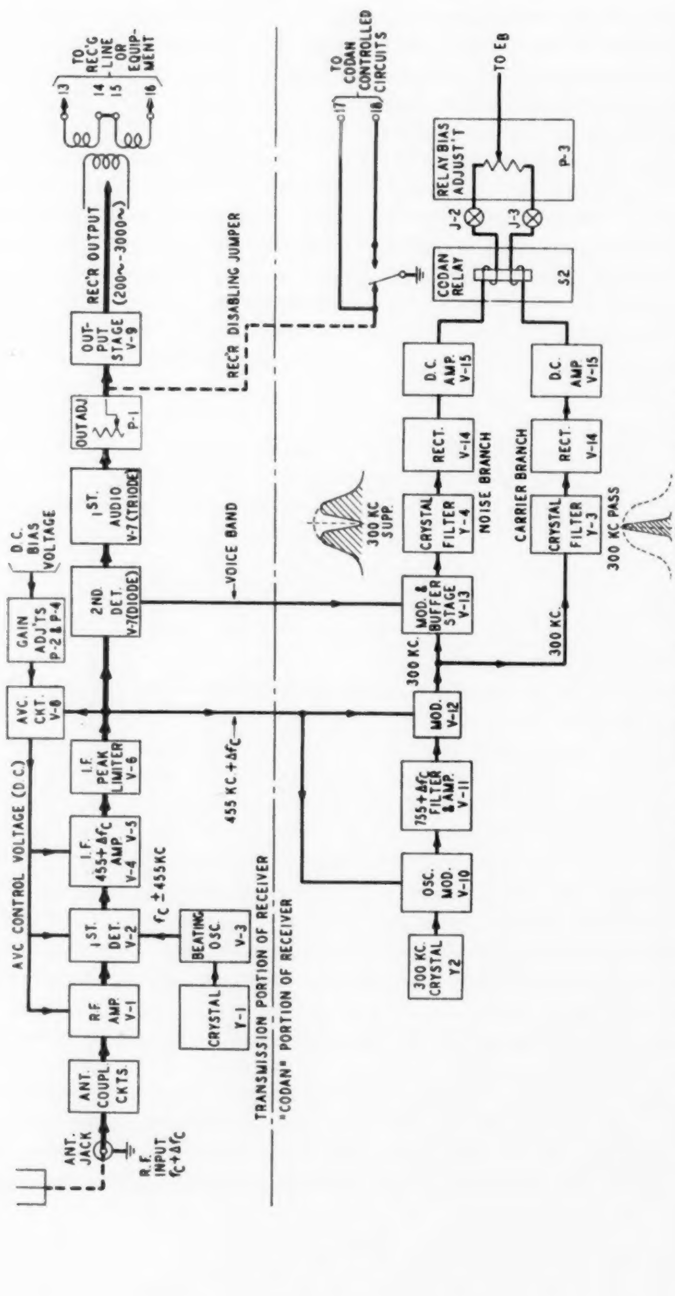


Fig. 9—Codan (carrier operated device anti-noise).

the delay circuits of the Vodas are unnecessary. This arrangement is finding increased application in ship-shore terminals.

Referring again to Fig. 5, note two squares in the first diagram labeled "Compandor." [4] Each of these squares has part of the name dotted to indicate that the two circuits are different but together form the entire Compandor. The Compandor is another device to assist in making signals more intelligible in the presence of noise at the receiver. It accomplishes this by the peculiar method of distorting the signal going out and then restoring it at the receiver. The reason for such a device and its mode of operation are as follows. Ordinary speech contains loud as well as weak signals. Most of the consonants and some of the vowels do not contain much energy. They therefore

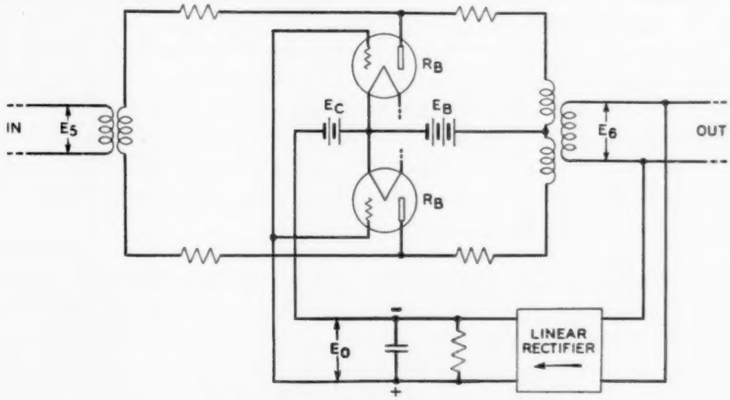


Fig. 10—Compressor part of the compandor.

will not modulate the transmitter fully and are the ones whose reception will be interfered with by noise at the receiver. The Compandor reduces this effect by making the weak parts of the transmitted signal larger than normal.

The part of the Compandor in the transmitter branch distorts the speech signal by reducing the energy variations between the loud and weak sounds a certain amount. It does not wipe out all variation as it is necessary to leave a certain variation which is made use of at the receiver to restore the original variation. The circuit used at the transmitter end is given in Fig. 10. It is called the Compressor. A speech signal comes in on the left-hand side and goes out on the right. Between the input and output circuits are connected two vacuum tubes. Although superficially this circuit looks as though these tubes are amplifiers, actually they are connected to absorb

energy. The operation therefore involves absorbing part of the energy in the louder signals and a lesser amount from the weaker signals so that the output contains speech which has been distorted in such a fashion that the variations in energy may be only one-tenth as much as they originally were. These two absorbing vacuum tubes are controlled by potential built up across a circuit containing capacitance and resistance. This circuit has a potential E_0 produced on it by a linear rectifier which secures its energy from the output circuit. The strong signals appearing in the output produce a larger voltage on the resistance-condenser combination, thereby causing the grids of the two vacuum tubes to be more positive than with weak signals or no signals, and the two tubes, then acting as conductive resistances, reduce the intensity of the loud signals. The resistance-condenser combination is so proportioned that the charge on the condenser rises and falls with syllabic frequency. It must not have such a short time constant as to wipe out individual cycles. It is to operate upon groups of cycles only. With this Compressor between the telephone line and the transmitter, and the amplifiers properly adjusted, the transmitter can still be fully modulated with the louder sounds in the voice but it will be modulated very much more than it would normally be by the weaker sounds in the voice.

At the receiving end the signal delivered by the receiver and transmitted towards the telephone line will be the same distorted signal which modulated the transmitter. Such a distorted signal, although scarcely discernible from the original, is not in all situations the desirable one to put upon a telephone line, so there are reasons for restoring this distorted signal to its original form. This distorted signal contains the weaker parts of speech amplified many times with respect to what would occur without the Compandor and therefore these weaker parts of speech will be many times above the noise which would have interfered with reception under ordinary conditions. This distorted speech now goes into the part of the Compandor in the receiving branch which is called the "Expandor," as shown in Fig. 11. The Expandor contains many elements similar to those in the Compressor but they are arranged in a slightly different form. Two vacuum tubes instead of absorbing energy are now used as amplifiers. The signal comes in on the left and goes out on the right but the output is not a true amplified picture of the input because as the signal goes through the amplifier the amplification of the two tubes is varied so as to restore the original signal. To do this use is made of the remaining amplitude variations within the signal to operate a linear rectifier and put a variable voltage E_x upon similar condenser

and resistance, and by virtue of connections to vary the amplification of the two amplifier tubes. The louder signals therefore put more positive bias on the grids of the amplifiers than the weaker ones and the amplifier tubes will amplify the stronger tones more than the weaker. There is thus delivered to the output the original signal but with very much improved signal-to-noise ratios.

In the use of the Compandor on circuits having noise it has been found possible to produce signal-to-noise improvements as high as 30 db. Average improvements are 15 to 20 db. The improvement depends upon the amount of noise present.

Returning again to Fig. 5, there are certain elements labeled "Calling." [3] The particular configuration indicated with calling devices attached to the telephone line in the first diagram and to a receiver

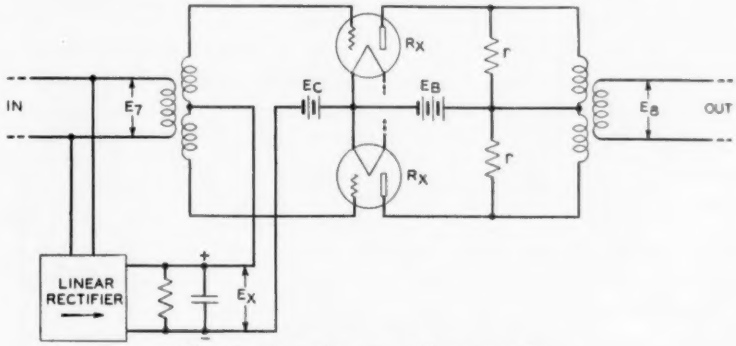


Fig. 11—Expandor part of the compandor.

in the second diagram are the particular calling arrangements used for communicating with boats at sea. The fishing boats, such as mentioned previously, and private yachts, do not find it expedient to have an operator listening at all times for calls nor do they like to have a loud speaker operating continuously delivering all and sundry communications to the people on a boat. It is desirable that means be provided so that the boat may be called by having a calling mechanism available. The system is indicated in Fig. 12. In this system the receiver on the boat must be operated continuously and must be connected to the selector and bell circuit so that whenever the correct calling signal comes in the bell will ring. At the shore station the calling is accomplished by sending out certain combinations of 600 cycles and 1500 cycles as indicated in this figure, the various combinations being chosen by the telephone dial which actuates a relay switching one or the other audio frequency onto the transmitter.

A certain combination which consists of certain sequences of the 600 and 1500-cycle tones is assigned to each boat. The 600 and 1500-cycle tones selected by band-pass filters are rectified and actuate a polar relay which then delivers to the part marked "selector" signals corresponding to those made by the telephone dial. The selector is a standard train dispatcher selector which can be set for various combinations of signals and when the correct combination arrives it will close a switch and ring the bell.

For calling the shore operator from the ship, the Codan mentioned previously connects the receiver to the line and also operates a relay to light the shore operator's switchboard light whenever a boat operator starts his transmitter and puts on his carrier.

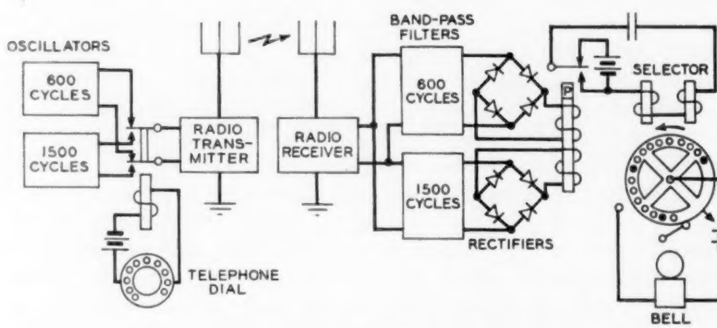


Fig. 12—Calling system for boats.

These two methods of calling in the two directions are not the only ones which are used on radio links. For most transoceanic and service to the large liners, prearranged schedules or continuous watch allow of calling by voice. In the Green Harbor-Provincetown circuit of Fig. 3 and some transoceanic circuits calling is accomplished by transmitting 1000 cycles interrupted 20 times a second which is a standard means of ringing over telephone lines.

Referring again to Fig. 5, note the third diagram and the element marked "SSBTR" [5] which means single sideband transmitter. Single sideband transmission has been used on transoceanic radio telephone circuits since the first circuit was opened. It is not used in broadcasting, at least in this country, although it has been proposed a number of times. Single sideband communication was first used on wire line carrier circuits with their inception in 1918. It was utilized when the long-wave radio circuit was tested experimentally in 1923

and opened in 1927. Single sideband has since been applied on short wave circuits to London and to Honolulu.

Single sideband has the theoretical advantage of 9 db in signal-to-noise ratio over double sideband with carrier transmission; 6 db is secured from the utilization of all of the energy in one sideband and 3 db comes from reduced noise in the reduced band width of the receiver. Tests have shown that the 9 db signal-to-noise improvement is secured in practice.

The application of single sideband to the short-wave circuits encountered a number of difficulties. One of the bigger difficulties was that of resupplying the eliminated carrier. In order that speech received over single sideband circuits be truly normal and be recognizable as the voice of the talker, the carrier must be resupplied within 20 cycles. If it is supplied more than 20 cycles out of position the speech will be intelligible in varying degrees but it is impossible to recognize the voice even of one's best friend. To resupply the carrier within 20 cycles when the radio frequency is 20 MC means that the transmitter and the beating oscillator at the receiver individually should not vary more than 10 cycles, and 10 cycles out of 20 MC is one part in 2 million. The frequency of either oscillator must therefore remain constant to better than one part in 2 million if voices are to be recognized. It is of course possible to build oscillators which are more stable than this. However, such oscillators at present appear to be in the laboratory rather than the commercial class and so it has been found desirable to adopt a different means for maintaining this frequency of the resupplied carrier at the right value. This is accomplished by transmitting a small part of the original carrier and then at the receiver use this small or vestigial carrier as it is known to actuate a mechanism which will supply a local frequency exactly in synchronism with it. The resupplied carrier can therefore be maintained well within one part in 2 million, in fact it can be maintained within one cycle in 20 megacycles.

In producing a single sideband other difficulties are encountered. In the present state of the art it is difficult to eliminate one sideband and leave the other, except at relatively low frequencies. In the long-wave transoceanic circuits which operate on 60 to 70 kc the elimination occurs at 30 kc and a second modulation shifts the remaining sideband to the desired position. This gives good selection of the desired sideband and provides flexibility in the final positioning of the sideband.

In operating at high frequencies, which may be as high as 22 MC, it has been found desirable to reach the desired point not with two steps in modulation but with three steps. This is indicated in Fig. 13.

Signals come in on input circuit A to modulator 1A which modulates 125 kc from which crystal filter A selects the upper sideband. This upper sideband then goes into modulator 2 where it modulates 2500 kc. The sideband in this case will be located more than 125 kc away from the carrier and can be selected by a relatively inexpensive filter which now delivers the single sideband in position 2625.1 to 2631 kc to modulator 3. In modulator 3 this sideband modulates a suitable

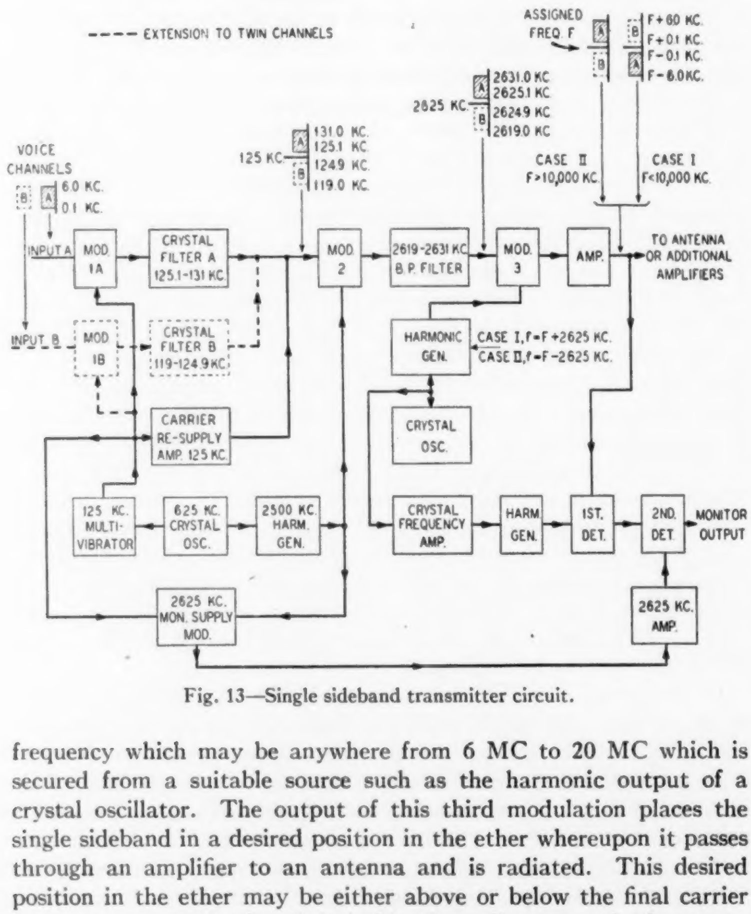


Fig. 13—Single sideband transmitter circuit.

frequency which may be anywhere from 6 MC to 20 MC which is secured from a suitable source such as the harmonic output of a crystal oscillator. The output of this third modulation places the single sideband in a desired position in the ether whereupon it passes through an amplifier to an antenna and is radiated. This desired position in the ether may be either above or below the final carrier frequency as may be found desirable, depending upon the frequency range in which operation is to occur at that particular hour.

Following the success in applying single sideband to one of the short-wave channels consideration has been given to utilizing the position of the vacated sideband or the contiguous position for a second channel. The resultant method is called "twin channel single sideband." Referring to the diagram in the upper right-hand corner of Fig. 13, Channel *A* is the single sideband which has just been discussed while the second single sideband is now placed in *B* position so as to give two separate conversations with the same transmitter. This immediately brings to mind the question, "Is the advantage of single side band lost by using the two sideband positions for two separate channels?" Odd as it may seem it is only slightly affected. If those two sideband positions are used for a single channel the frequencies in the two sidebands appear simultaneously, and in corresponding positions. The transmitter must handle both simultaneously and the receiver must be broad enough to receive both bands. However, when two separate conversations are placed in the two separate sideband positions similar frequencies do not appear simultaneously in both bands except at such remotely occasional times that their mutual interference is small or not noticeable, and at the same time the receiver for each channel is tuned for only one sideband, thereby keeping down the noise. By using the two positions for two separate channels it is possible to get on a statistical basis two single sideband circuits each 8 db better in signal-to-noise ratio than would be obtained using the same two sideband positions for one channel alone. It produces a remarkable increase in efficiency of use of a circuit.

Now it also happens that addition of the second channel requires a surprisingly small amount of apparatus. Looking at this same figure, Channel *B* is indicated on the left providing input *B* to modulator 1*B*. This modulator modulates the same frequency as modulator 1*A* but crystal filter *B* selects the lower sideband in this case, which sideband is now delivered to modulator 2 along with that from filter *A*. These two parts are all the apparatus necessary to add to this transmitter to convert it from one channel to two channels. It is thus to be observed that by suitable application of single sideband to the short-wave channels it has been possible to multiply their number by two and increase each one 8 db in signal-to-noise ratio. This twin channel single sideband has been applied to two of the three short-wave transoceanic circuits, to the San Francisco-Honolulu circuit, and undoubtedly will be applied to other circuits in the future.

Referring again to Fig. 5, there will be seen an element marked "Directive Antenna." Directive antennas have been used very little in broadcasting. They are coming into greater use with short-wave

broadcasting and with efforts to produce less interference between stations in the United States operating on similar wave-lengths, but their use has been much smaller than their use in radio links of the telephone system. Directive antennas are of great importance in telephone links for the reason that in operating over great distance, where weak signals must be received all or most of the time, much power may be saved if directive antennas at the transmitter are used

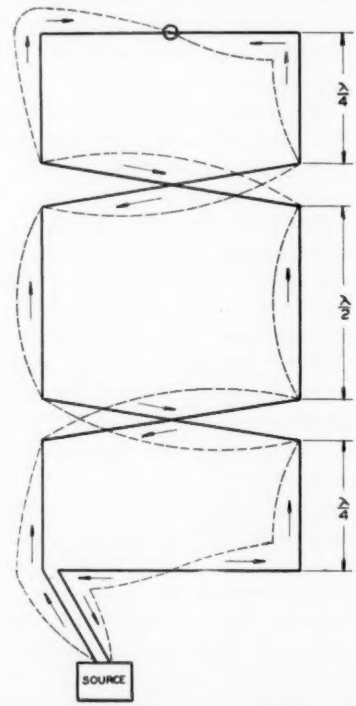


Fig. 14—Element of the Sterba directive antenna.

to send as much energy as possible in the desired direction, and used at the receiver to exclude as much noise from undesired directions as can be done.

In connection with the directive antennas used in the telephone system, quite a variety have been developed. Figure 14 indicates the principle of what is known as the Sterba [6] antenna, invented by the late Mr. E. J. Sterba. This is an elemental section, the complete antenna being built up of a number of elements of this kind. The

element consists of a conductor, whose length is an integral number of half wave-lengths, bent in such a way that the currents in all of the vertical elements will be in phase while in the horizontal elements they will largely balance out. The vertical elements are all placed in the same plane and the horizontal elements depart from that plane only enough for crossing without short-circuiting. With this arrangement the energy radiated in a direction perpendicular to the paper will be a maximum. It will be a minimum in the plane of the paper in side or vertical directions. Any suitable number of these elements may be arranged in the same plane and connected together so as to increase the energy in the desired direction. Inasmuch as the length of the wire and the configuration into which it is bent are associated with the frequency, an antenna constructed for one frequency is not usable at another. This is true in any antenna where standing waves exist.

In the antennas of this type which have been used on our short-wave circuits for operating across the Atlantic, as many as 8 elements were connected in parallel in the same plane so as to radiate in the desired direction producing a sharp directivity pattern. At the same time another set of 8 elements were located one-half wave-length away parallel to the first as indicated in Fig. 15 so as to eliminate the radiation in one of the two directions perpendicular to the screen. This causes all the energy to be radiated in the desired direction. For operating one radio transmitter on a transoceanic circuit it is necessary to have three or four antennas for each transmitter, one for each wave-length. These antennas were strung between towers. Figure 16 shows the antennas as used formerly at Lawrenceville, New Jersey. One antenna occupied two inter-tower spacings. The direction of transmission is perpendicular to the line of the towers.

At the receiving end directive antennas are also used. An elemental picture is shown in Fig. 17 of the receiving antenna devised for this purpose by Mr. E. Bruce.[7] The third diagram shows the shape into which a long conductor is bent and the arrows show the instantaneous directions of current flow at a moment of maximum. It is observed here also that in all the vertical elements the currents flow in the same direction while in the horizontal elements enough flows in the two directions so the effect is neutralized. This antenna will therefore also receive or transmit in the directions perpendicular to the plane of the conductor and not in directions within the plane. This type of antenna can also be constructed with a reflector behind it to reduce the direction of transmission or reception to a single direction.

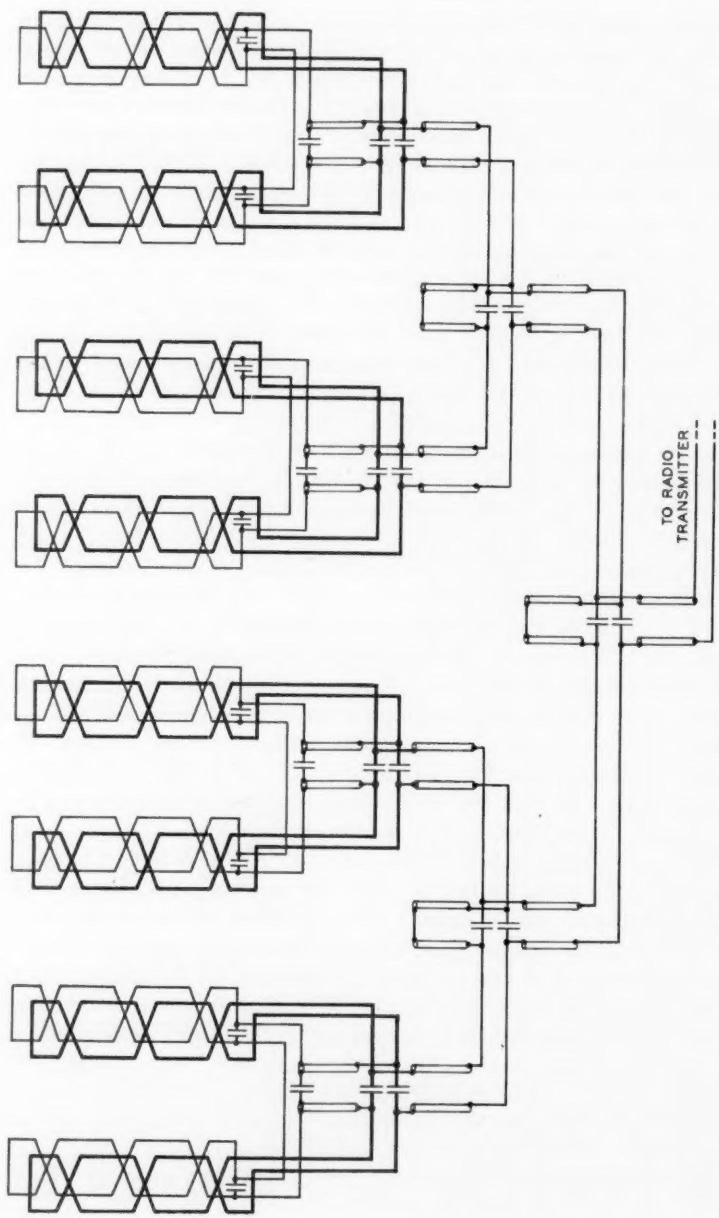


Fig. 15—A complete Sterba antenna with reflectors for one wave-length.

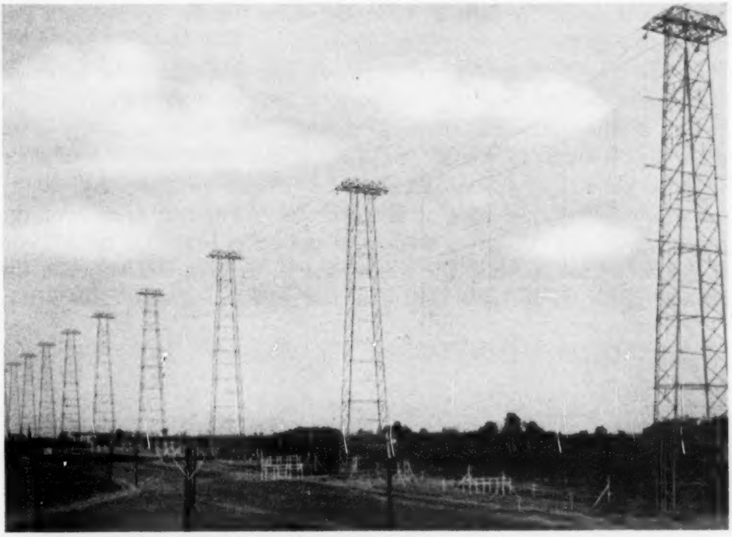


Fig. 16—Towers at Lawrenceville, New Jersey, used to support a number of Sterba antennas for transoceanic communication. These antennas have since been superseded by rhombic antennas.

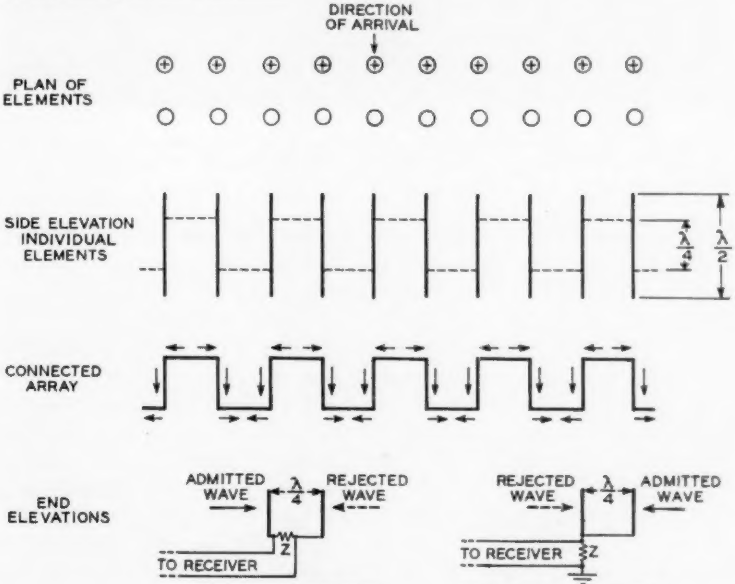


Fig. 17—Bruce antenna array.

Such an antenna as used at Netcong, New Jersey, for reception is shown in Fig. 18.

In the course of time further improvements in directive antennas have been made. One improvement is indicated in Fig. 19 and is known as the "rhombic antenna" [8] because it consists of a wire or wires supported by four poles in the shape of a rhombus elevated some distance above the ground and parallel to the ground. As usually used two contiguous sides of the rhombus form one branch of the antenna and the other two sides form the other branch. At one end is connected the receiver (or transmitter). At the end opposite the connection to the receiver are connected resistances of suitable value.

Fig. 18—Bruce receiving antenna at Netcong, New Jersey, used for transoceanic communication. This type of antenna has been superseded by the rhombic type at Netcong.

This antenna differs radically from the previous ones and most other directive antennas in one respect and that is it does not usually operate with a standing wave thereon. The purpose of the resistors is to absorb all the energy reaching such resistors. This antenna when arranged as in Fig. 19 receives from the right. The energy strikes the wires and is thence transmitted to the receivers. Energy coming from the left-hand side travels away from the receivers and when it strikes the resistors is absorbed. If the resistors are not used and the two terminals are either connected together or kept insulated, energy reaching this end will be reflected, in which case this antenna will operate with a standing wave thereon and will receive or transmit from either the right or the left Inasmuch as this antenna as preferably used is unidirectional and does not operate with a standing

wave, a single antenna may be used for a number of wave-lengths without readjustment.

When the rhombic antenna is used for a number of frequencies without change in size or form, the directivity is different for each frequency. The maximum directivity for the higher frequencies will be a lower angle than for the lower frequencies. This works in very well for long distance operation since the angles at which the high and low frequencies come in tend to agree with this characteristic of the antenna.

The rhombic antenna has come into extensive use for transoceanic short-wave links during the last few years. Due to its multi-wave-

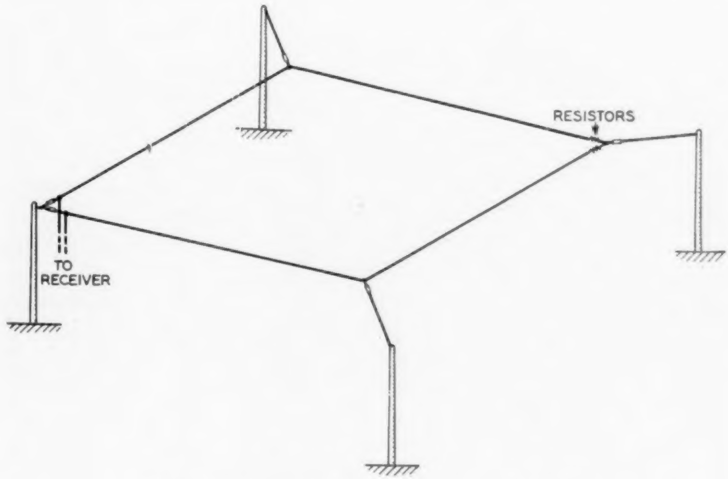


Fig. 19—Rhombic antenna invented by E. Bruce.

length characteristics it is replacing the other types of antennas described previously. The initial cost is much less than the type supported on steel towers, the land required is less, and the upkeep is also less.

In Fig. 20 is shown a photograph of another type of directive antenna. Two antennas are indicated, one for transmitting and one for receiving. These antennas are known as "pine tree" [9] antennas because of the connections of the radiators to a transmission line passing up from below. This particular antenna is for ultra-short-wave operation around 60 MC and is one end of the Provincetown-Green Harbor circuit. Each antenna contains 8 radiators in a plane

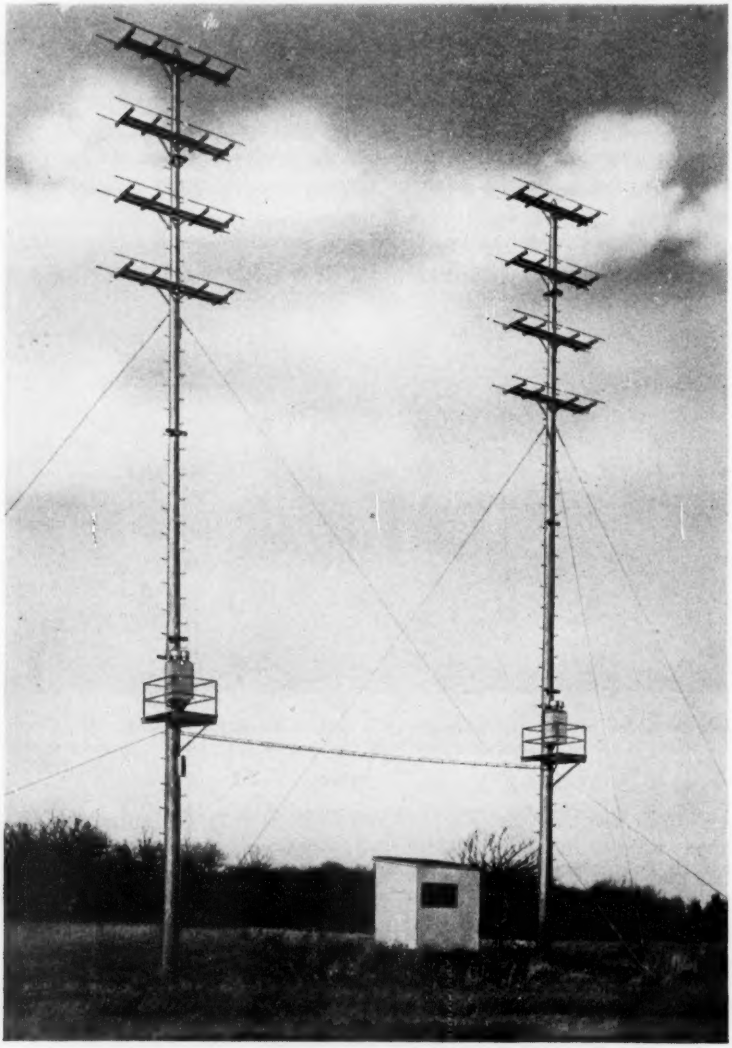


Fig. 20—Pine tree antennas used on Provincetown-Green Harbor ultra-short-wave circuit.

and 8 reflectors in a parallel plane. This type lends itself very well to construction for ultra-short-wave operation. The directivity of any antenna is a function of its breadth measured in wave-lengths and when the wave-length is short it is more economical and easier to construct the antenna with the larger dimensions up and down the pole or at right angles to the pole than to stretch it along the ground.

Figure 21 indicates one of the first directive antennas to be used. This is known as the "Beverage" [10] antenna and is used on long-wave telephone circuits which operate between 60 and 70 kc. Such an antenna is located at Houlton, Maine, for receiving from England. The antenna structure appearing in this diagram may be several miles in length. It receives best from the upper right-hand direction, as indicated by the arrow. The incoming wave produces currents in the antenna which experience gradual build-up along the line to the receiving end where they operate the receiver and deliver the signal to the telephone line. This type of antenna has a horizontal directional pattern as indicated in the figure. Its maximum direction as used is northeast, as that is the direction signals arrive from England. It also happens that more static and strays reach this part of the country from the southwest than from other directions and with the antenna so oriented there is what is sometimes called a "blind eye" faced in the southwest direction so as to receive a minimum of interference from that direction.

Some of the difficulties involved in transoceanic short-wave reception may be explained by reference to Fig. 22; this shows a diagram of the earth and the ionized region of the atmosphere called the "ionosphere." Signals from the transmitting station in England may reach the receiving station in the United States by more than one path. One path indicated has two reflections from the ionosphere and the other has three reflections. Careful measurements on this diagram indicate that these two paths are not equal in length, with the result that signals received in the United States from across the ocean coming over the two paths may be out of phase. Not only may there be two paths but sometimes there are three, four or more so that the interference caused by signals coming over the several paths can give rise to bad fading and distortion. The lower diagram in this figure shows the vertical directive pattern of an ordinary directive antenna. It shows this directive pattern to be large enough to receive simultaneously both incoming signal components from the two paths. If it is desired to eliminate the undesirable effects produced by the two signals coming in out of phase, one should be eliminated. This can be done provided an antenna is constructed having a sharp directive

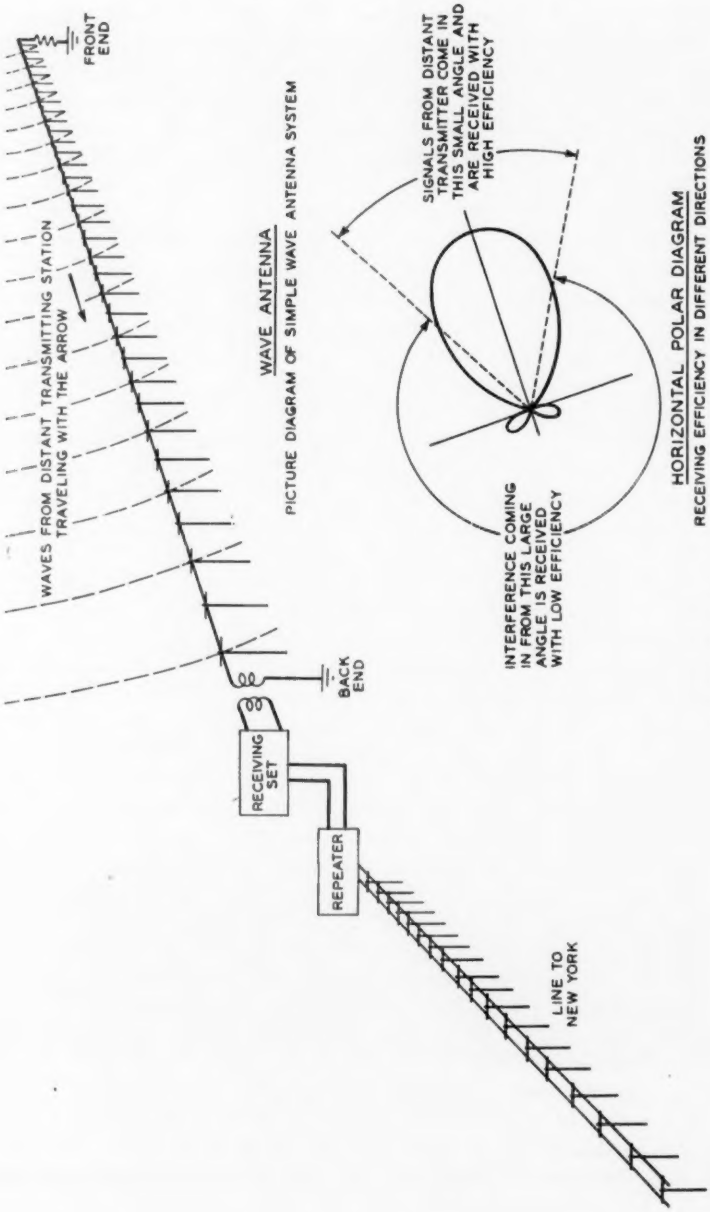
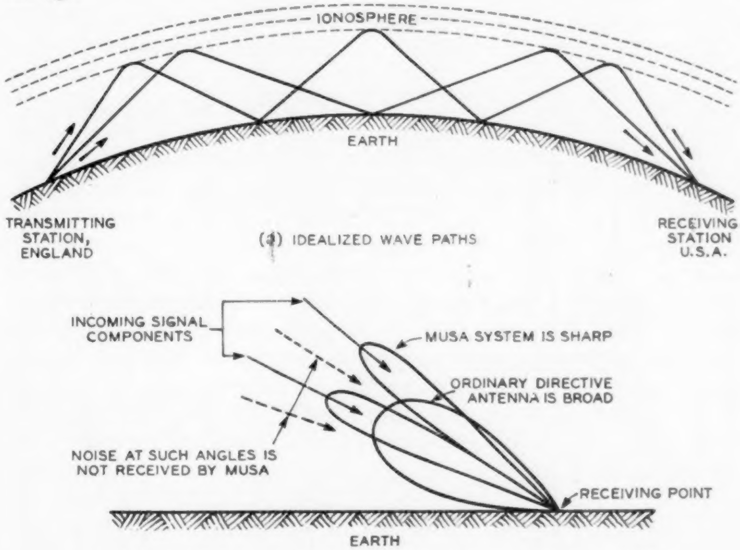


Fig. 21—Beverage antenna.

pattern as indicated in the same sketch and such a directive pattern is produced by an antenna system called the "Musa." [11] The name "Musa" comes from the initials of the words "multiple unit steerable antenna." The Musa is one of the latest developments in directive antennas and possesses not only the important characteristic of a very sharp directivity pattern but also is steerable so it may be altered to receive a desired component, and if such desired component changes its angle of arrival alteration may be made to accommodate such change.



(b) VERTICAL PLANE DIRECTIVE PATTERNS (NOT DRAWN TO SCALE)

Fig. 22—Paths of short waves over long distances as determined by the ionosphere.

Figure 23 indicates the elements of the Musa. It consists of a row of rhombic antennas lined up in the direction from which signals are expected. Each rhombic antenna is connected by a transmission line to a phase shifter and the outputs of the phase shifters are connected to a receiver. These phase shifters may be so adjusted as to cause any desirable phase additions from the separate antennas. By changing the adjustments on the respective phase shifters the direction of reception may be altered. In this diagram a row of antennas is shown as connected through phase shifters to receiving branch A with the phase shifters adjusted to produce a directive pattern as indicated in the neighboring diagram marked branch A and drawn dotted. Into branch A will therefore come the signals which arrive from one trans-

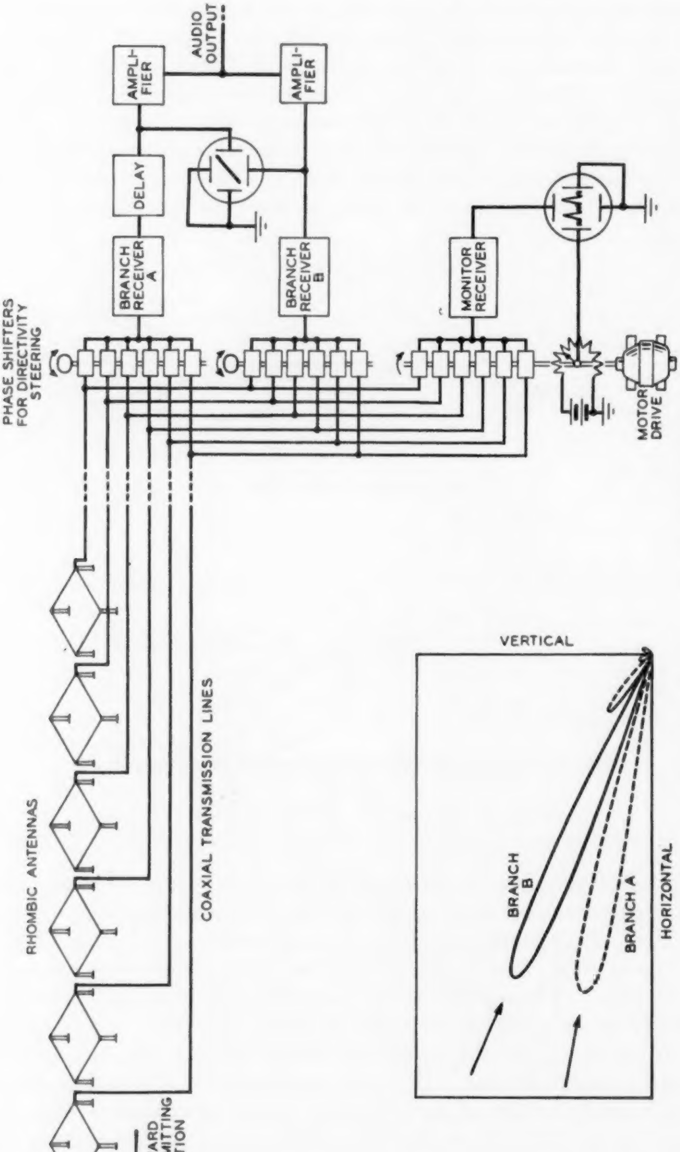


Fig. 23—Schematic of musa (multiple unit steerable antenna).

oceanic path at the lower angle. Beneath this row of phase shifters is indicated a second set connected in parallel with the first set. The second set is adjusted somewhat differently so that the receiver in branch *B* will receive signals over the path with the higher angle as indicated by the solid line directive pattern in the neighboring diagram. The two separate signals may therefore be separately received. Since, however, they do not arrive at identical times they cannot be added without producing distortion. By interposing a delay in the output branch of the one arriving earlier, which in this case is branch *A*, they may be added directly into a single audio output. In this case if one branch fades the other will continue to receive the signal, and while both are receiving a better signal-to-noise ratio obtains.

A cathode ray oscillograph is connected on the outputs of these receivers as indicated in the diagram so as to indicate when the delay is satisfactory for adding the two signals in phase.

As mentioned previously the direction from which a signal comes may change from time to time. In this case a fixed adjustment of the phase shifters may allow the signal to disappear. The operator needs to be ready to change the phase shifters when necessary but it is also desirable that he change them in the desired direction without interfering with the conversation arriving over the circuit, which means he should be able to adjust them without using cut-and-try methods of adjustment. To accomplish this a third set of phase shifters is provided which are continually driven in rotation by a motor so that as they rotate they cause the direction of the signal received by a monitoring receiver to shift its angle of reception over the entire angle range at which signals might arrive. The output of this receiver is connected to two plates of a cathode ray oscillograph while to the other two plates are connected potentials from a rheostat on the phase shifter drive shaft so that the potential will be indicative of the position of the phase shifter. If the phase shifter continually rotates the varying signals will cause the varying deflection in the vertical direction and so draw a pattern as indicated in Fig. 23, which will show immediately by suitable calibration the angle between the two different signals. The operator may thereby adjust his phase shifters by calibration and without cut-and-try tuning.

Inasmuch as the rhombic antenna does not operate with standing waves and may be used for a variety of frequencies it may be used simultaneously for this variety of frequencies. The transmission lines from the rhombic antennas to the phase shifters also operate without standing waves so they likewise may operate simultaneously at a variety of frequencies. It is thus possible by connecting other phase shifters and other receivers to utilize a single row of rhombic antennas

to receive simultaneously a number of radio signals on different frequencies. Each one of these receivers with phase shifters will provide its own directivity pattern and may be adjusted independently of the others.

A Musa system has been constructed for transoceanic reception on short waves at Manahawken, New Jersey, and is now in operation. In this system 16 rhombic antennas are placed in a row approximately a mile and a half long.

The radio terminals for all services have not attained standardized final forms but are in a slow state of flux as better circuits and methods for handling existing problems are devised. The circuits and devices indicated in the figures in principle or in detail are not the only ones that have been tried or used, but represent steps that at one time or another were considered to be advancements suitable to be put into use while the attentions of the development engineers were directed toward more pressing problems. These various devices have augmented the reliability of radio circuits enormously. Distances covered have been enlarged. To accomplish similar results by power increase alone would have in most cases rendered it uneconomic to construct and operate the radio systems. In each case peculiarities either in speech, in radio circuits, or in static and noise characteristics are taken advantage of in making a design to aid the signal and reduce the effect of noise. It is believed that the limit has not been reached but that further improvements and other devices will in due time give increased reliability.

References

1. "The Vodas," S. B. Wright, *Electrical Engineering*, August 1937, or *Bell System Technical Journal*, October 1937.
2. "A Vogad for Radiotelephone Circuits," S. B. Wright, S. Doba and A. C. Dickieson, *I.R.E. Proceedings*, April 1939.
3. "A Radio Telephone System for Harbor and Coastal Service," C. N. Anderson and H. M. Pruden, *I.R.E. Proceedings*, April 1939.
4. "The Compandor, an Aid Against Static in Radio Telephony," R. C. Mathes and S. B. Wright, *Electrical Engineering*, June 1934, and *Bell System Technical Journal*, July 1934.
5. "A Short Wave Single Sideband Radio Telephone System," A. A. Oswald, *I.R.E. Proceedings*, December 1938.
6. "Theoretical and Practical Aspects of Directional Transmitting Systems," E. J. Sterba, *I.R.E. Proceedings*, July 1931.
7. "Transoceanic Telephone Service," A. A. Oswald, *Bell System Technical Journal*, vol. IX, p. 270.
8. "Developments in Short Wave Antennas," E. Bruce, *I.R.E. Proceedings*, August 1931, and "Horizontal Rhombic Antennas," E. Bruce, A. C. Beck and L. R. Lowry, *I.R.E. Proceedings*, January 1935.
9. "An Unattended Ultra-Short-Wave Radiotelephone System," N. F. Schlaack and F. A. Polkinghorn, *I.R.E. Proceedings*, November 1935.
10. "The Wave Antenna," H. H. Beverage, C. W. Rice and E. W. Kellog, *Trans. A. I. E. E.*, vol. 42, p. 215, 1923.
11. "A Multiple Unit Steerable Antenna for Short-Wave Reception," H. T. Friis and C. B. Feldman, *I.R.E. Proceedings*, July 1937.

Abstracts of Technical Articles by Bell System Authors

Lodgepole Pine Poles—Full Length Treatment Under Pressure—Butt Treatment in Open Tanks.[1] C. H. Amadon. Lodgepole pine ("Pinus contorta") forms extensive forests in Colorado, Wyoming, Idaho and Western Montana. The timber has been used widely for mine props, railway ties and telephone and telegraph poles by the various industries in the general region in which it grows. Like most of the pines, lodgepole pine in its natural state is classed as a non-durable wood in contact with the soil, and where a relatively long service life is desired the timber has been treated with creosote or some other wood preservative.

Lodgepole pine poles are exceptionally straight, free from knots of objectionable size, fairly soft and when well seasoned weigh about 30 lb. per cu. ft.

The purpose of this paper is to present information on the behavior, under actual service conditions, of lodgepole pine poles that had been pressure-treated in closed cylinders or butt-treated in open tanks, and to describe the development of a process for the pressure treatment of lodgepole pine poles to meet specific penetration and low retention specification.

Sound Measurement Objectives and Sound Level Meter Performance.[2] J. M. Barstow. The standardization of sound level meters is shown to have improved conditions in the field of sound measurement, although several characteristics thought to be desirable in visual indicating sound measuring devices are not fully realized in instruments conforming to the present standards. The extent to which certain sound measurement objectives have been realized in present sound level meters is discussed. Further work will undoubtedly be necessary before some of these objectives may be more completely realized. Present indications are that sound level meter limitations in regard to the approximation of sound jury loudness levels will be difficult to remove and at the same time retain reasonable apparatus simplicity. Some consideration is given to possible courses of action in regard to such limitations.

Coordination of Power and Communication Circuits for Low-Frequency Induction.[3] J. O'R. Coleman and H. M. Trueblood. Where power

[1] *Proc. Amer. Wood-Preservers' Assoc.*, 1940.
[2] *Jour. Acous. Soc. Amer.*, July 1940.
[3] *Electrical Engineering*, July 1940.

and communication facilities are in proximity, electromagnetic induction from the power system may cause disturbances in the communication system. The avoidance or minimizing of such disturbances, with due regard to the service and other needs of both systems, is a problem of coordination, which is conveniently divided into two parts, one dealing with low-frequency inductive coordination and the other with noise-frequency coordination.

The present paper undertakes a general examination of the problem of low-frequency inductive coordination in the light of developments during the past decade. The situation as it existed at the beginning of the decade is to be found well set forth in a paper presented in 1931 at the A.I.E.E. winter convention by R. N. Conwell and H. S. Warren. The present paper, like its predecessor, derives from the work of the Joint Subcommittee on Development and Research of the Edison Electric Institute and the Bell System. It is largely concerned with induction from currents due to power system ground faults and the transients which accompany such faults. It gives relatively little attention to continuous low-frequency effects since, up to the present at least, such effects have not been a primary concern in the low-frequency coordination of commercial power circuits and Bell System communication circuits.

A further object of the paper is to outline the various factors that require consideration in practical situations and to discuss their significance under present-day conditions. To provide necessary background for this, recapitulations of fundamentals are included at appropriate points. Detailed discussions necessarily omitted from the paper itself are to be found in the papers listed in the bibliography, many of which, particularly the Conwell-Warren paper, contain further references

Insulating Paper in the Telephone Industry.[4] J. M. FINCH. This article discusses briefly a few of the more important types of paper insulations used by the telephone industry, and shows the relation the manufacturing procedures bear to the initial properties, the permanence, and the uses of the product. Special emphasis is placed on chemical properties as criteria of permanence. The specification control of paper is discussed with emphasis on the simplification of chemical test methods and on minimizing the number of such tests. Finally, mention is made of some of the modified forms of cellulose, which possess insulating characteristics superior to paper and which are already replacing it for some uses.

[4] *Indus. and Engg. Chemistry,* August 1940.

Rectilinear Electron Flow in Beams.[5] J. R. PIERCE. Electrodes are devised by means of which rectilinear electron flow according to well-known space charge equations can be realized in beams surrounded by charge-free space. It is shown how these electrodes can be used in the design of electron guns having desirable characteristics.

High-Gain Amplifier for 150 Megacycles.[6] G. RODWIN and L. M. KLENK. An ultra-high-frequency amplifying system is described which operates at about 150 megacycles with an over-all gain of 114 decibels and transmitted band of over 2 megacycles. An output power of 2.5 watts is available with a signal-to-distortion ratio of 60 decibels. By a frequency-shifting modulator in the amplifier chain the input and output are made to differ by 10 megacycles. A filter-type circuit is used as the interstage coupling to give the necessary band width.

Room Noise at Subscribers' Telephone Locations.[7] D. F. SEACORD. The effect of room noise on the ability to hear speech is roughly equivalent to a partial deafening of the listener; hence the study of room noise conditions at telephone locations is of considerable interest to the telephone engineer since these conditions have an important bearing on the degree of satisfaction with which speech is received over a telephone connection. As a consequence, various studies of room noise have been made from time to time and information of increasing value has been obtained over a period of years with the development of improved measuring equipment and technique. This paper is based on the results of recent room noise surveys carried out in the Bell System and gives a broad picture of the magnitude of room noise at subscribers' telephone locations under present-day conditions. The data presented are a part of the information required in the work of devising and applying methods for taking into account the effects of room noise on telephone transmission in the design of the telephone plant.

Temperature Effects in Secondary Emission.[8] D. E. WOOLDRIDGE. Measurements have been made on the effects of temperature changes on the emission of secondary electrons from iron, nickel, cobalt, and molybdenum. Abrupt changes of one or two per cent were observed to accompany the $\alpha-\gamma$ transition of iron, while the hexagonal to face-centered cubic transformation of cobalt was accompanied by a change in secondary emission of only about 0.4 per cent. The magnetic trans-

[5] *Jour. Applied Physics*, August 1940.
[6] *Proc. I. R. E.*, June 1940.
[7] *Jour. Acous. Soc. Amer.*, July 1940.
[8] *Phys. Rev.*, August 15, 1940.

formation was found to alter the secondary emission coefficient of nickel by less than 0.3 per cent. The temperature coefficient of secondary emission, in the cases of nickel, cobalt, and molybdenum, was found to be much less than the volume coefficient of expansion of the metal. The smallness of the temperature coefficient and the effect of the magnetic transformation are shown to lend support to the view that the secondary electrons are scattered or "absorbed" by an excitation process similar to that whereby they are originally produced.

Contributors to this Issue

MILLARD W. BALDWIN, JR., E.E., Cornell, 1925; M.A., Columbia, 1928. Bell Telephone Laboratories, 1925–. Mr. Baldwin's work has been in the field of telephotography and television.

W. R. BENNETT, B.S., Oregon State College, 1925; A.M., Columbia University, 1928. Bell Telephone Laboratories, 1925–. Mr. Bennett has been engaged in the study of the electrical transmission problems of communication.

G. K. BURNS, S.B., S.M., Massachusetts Institute of Technology, 1935. Western Electric Company, 1935–1940; Bell Telephone Laboratories, 1940–. Mr. Burns has been engaged in the development of methods and facilities for the manufacture of crystal filters. At present he is on a temporary assignment with the Laboratories' filter design group.

HOMER DUDLEY, B.S. in Electrical Engineering, Pennsylvania State College, 1921; M.A., Columbia University, 1924. Bell Telephone Laboratories, 1921–. Mr. Dudley has worked on transmission problems. In recent years, the work has been in research on circuits for speech analysis and synthesis.

MARK B. GARDNER, B.S., Brigham Young University, 1930. Bell Telephone Laboratories, 1930–. Mr. Gardner has been engaged in various studies in physiological acoustics.

RAYMOND A. HEISING, E.E., University of North Dakota, 1912; M.S., University of Wisconsin, 1914. Western Electric Company, 1914–25; Bell Telephone Laboratories, 1925–. Mr. Heising has been engaged in research and development on all phases of radio and is best known for his work on radio transmitters.

H. C. MONTGOMERY, A.B., University of Southern California, 1929; M.A., Columbia University, 1933. Bell Telephone Laboratories, 1929–. Engaged at first in studies of hearing acuity and related problems in physiological acoustics, Mr. Montgomery has been occupied more recently with the study and analysis of speech.

J. C. STEINBERG, B.Sc., M.Sc., Coe College, 1916, 1917. U. S. Air Service, 1917–19. Ph.D., Iowa University, 1922. Engineering Department, Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Dr. Steinberg's work since coming with the Bell System has related largely to speech and hearing.

VOLUME XIX OCTOBER, 1940 NUMBER 4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50c per Copy $1.50 per Year

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are fifty cents each.
The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.